Exponential Synchronization for Fractional-order Time-delayed Memristive Neural Networks

Ding Dawei[1], Zhang Yaqin[2] and Wang Nian[3*]
School of Electronics and Information Engineering, Anhui University
Hefei, 230601, China
[1]e-mail: dwding@ahu.edu.cn
[2]e-mail: 1970359758@qq.com
[3*]e-mail: wn_xlb@ahu.edu.cn

***Abstract*—Considering the fact that the exponential synchronization of neural networks has been widely used in theoretical research and practical application of many scientific fields, and there are a few researches about the exponential synchronization of fractional-order memristor-based neural networks (FMNN). This paper concentrates on the FMNN with time-varying delays and investigates its exponential synchronization. A simple linear error feedback controller is applied to compel the response system to synchronize with the drive system. Combining the theories of differential inclusions and set valued maps, a new sufficient condition concerning exponential synchronization is obtained based on comparison principle rather than the traditional Lyapunov theory. The obtained results extend exponential synchronization of integer-order system to fractional-order memristor-based neural networks with time-varying delays. Finally, some numerical examples are used to demonstrate the effectiveness and correctness of the main results.**

Keywords-Exponential Synchronization; Memristor-based Neural Networks; Fractional-order; Linear Error Feedback Control; Time-varying Delays.

I. INTRODUCTION

Chua already supposed the existence of memristor in 1971 [1], however, the practical device of memristor in electronics is obtained in [2] until 2008. In addition to the existing three kinds of circuit elements, memristor is regarded as the fourth basic circuit element and is defined by a nonlinear charge-flux characteristic. As everyone knows, resistors can be used to work as connection weights so that it can emulate the synapses in artificial neural networks. However, in the neural networks of biological individual, long-term memories is essential in the synapses among neurons, but for the general resistors, it is impossible to have the function of memory. Recently, due to the memory characteristics of memristor, memristor can replace the resistor to develop a new neural networks that is memristor-based neural networks (MNN) [3-6].

In recent years, more and more attentions have been put on the dynamical analysis of memristor-based neural networks, such as the investigation of stability [7-10], periodicity [11-13], system synchronization [14-22], passivity analysis [23], dissipativity [24-25] and attractivity [26]. Particularly, the stability and synchronization of MNN has been widely studied in [27-30]. In fact, synchronization means the dynamics of nodes share the common time-spatial property. Therefore we can understand an unknown dynamical system by achieving the synchronization with the well-known dynamical systems [18]. Moreover, in the transmission of digital signals, communication will become security, reliable and secrecy by achieving synchronization between the various systems. Therefore, the synchronization of MNN is still worth further research.

Moreover, the fractional-order models can better describe the memory and genetic properties of various

materials and process, so the fractional-order models have received a lot of research attentions than integer-order models. In recent years, with the improvement of fractional-order differential calculus and fractional-order differential equations, it is easy to model and analyze practical problems [31, 32]. Therefore, there have been a lot of researches about the dynamical analysis and synchronization of fractional-order memristor-based neural networks (FMNN) [34-39]. Finite-time synchronization, hybrid projective synchronization and adaptive synchronization of FMNN have all been researched [34-36]. However, there are only a very few research results on exponential synchronization of FMNN. In fact, the exponential synchronization of neural networks has been widely used in the theoretical research and practical application of many scientific fields, for example, associative memory, ecological system, combinatorial optimization, military field, artificial intelligence system and so on [40-43]. So the exponential synchronization of FMNN is still worth further studying as it is a significant academic problem.

On the other hand, the stability and synchronization of FMNN without time delay have been deeply studied such as in [33]. However, in hardware implementation of neural networks, time delay is unavoidable owing to the finite switching speeds of the amplifiers. And it will cause instability, oscillation and chaos phenomena of systems. So the investigation for stability and synchronization of FMNN cannot be independent on the time delay.

Motivated by the above discussion, this paper studies the exponential synchronization of FMNN with time-varying delays. The main contributions of this paper can be listed as follow. (1) This is the first attempt to achieve exponential synchronization of FMNN with time-varying delays by employing a simple linear error feedback controller. (2) The sufficient condition for exponential synchronization of FMNN with time delays is obtained based on comparison principle instead of the traditional Lyapunov theory. (3) Some previous research results of exponential synchronization for integer-order memristor-based system are the special cases of our results. Furthermore, some numerical examples are given to demonstrate the effectiveness and correctness of the main results.

The rest of this paper is organized as follows. Preliminaries including the introduction of Caputo fractional-order derivative, model description, assumptions, definitions and lemmas are presented in Section 2. Section 3 introduces the sufficient condition for exponential synchronization of the FMNN. In Section4, the numerical simulations are presented. Section5 gives the conclusion of this paper.

II. PRELIMINARIES

Compared to the integer-order derivatives, we know the distinct advantage of Caputo derivative is that it only requires initial conditions from the Laplace transform of fractional derivative, and it can represent well-understood features of physical situations and making it more applicable to real world problems [36]. So in the rest of this paper, we apply the Caputo fractional-order derivative for the fractional-order memristor-based neural networks (FMNN) and investigate the exponential synchronization of FMNN.

A. The Caputo fractional-order derivative

Definition1 [32] The Caputo fractional-order derivative is defined as follows:

$$D_t^q f(t) = \frac{1}{\Gamma(m-q)} \int_{t_0}^{t} \frac{f^{(m)}(\tau)}{(t-\tau)^{q-m+1}} d\tau, \tag{1}$$

where q is the order of fractional derivative, m is the first integer larger than q, $m-1 \le q < m, \Gamma(\cdot)$ is the Gamma function,

$$T(x) = \int_0^{\infty} t^{x-1} e^{-t} dt. \tag{2}$$

Particularly, when $0 < q < 1$,

$$D_t^q f(t) = \frac{1}{\Gamma(1-q)} \int_{t_0}^{t} \frac{f'(\tau)}{(t-t_0)^q} d\tau. \tag{3}$$

B. Model description

In this paper, referring to some relevant works on FMNN [35,36], we consider a class of FMNN with time-varying delays described by the following equation,

$$D^q x_i(t) = -c_i x_i(t) + \sum_{j=1}^{n} a_{ij}\left(x_j(t)\right) f_j\left(x_j(t)\right) + \sum_{j=1}^{n} b_{ij}\left(x_j\left(t-\tau_j(t)\right)\right) g_j\left(x_j\left(t-\tau_j(t)\right)\right) + I_i,$$
$$t \geq 0, i \in N.$$
$$a_{ij}\left(x_j(t)\right) = \frac{M_{ij}}{C_i} \times \delta_{ij}, \quad b_{ij}\left(x_j\left(t-\tau_j(t)\right)\right) = \frac{W_{ij}}{C_i} \times \delta_{ij}, \quad \delta_{ij} = \begin{cases} 1, & i \neq j, \\ -1, & i = j, \end{cases} \tag{4}$$

where $x_i(t)$ is the state variable of the i th neuron (the voltage of capacitor C_i), q is the order of fractional derivative, $c_i > 0$ is the self-regulating parameters of the neurons, $0 \leq \tau_j(t) \leq \tau$ and (τ is a constant) represents the transmission time-varying delay. $f_j, g_j : R \to R$ are feedback functions without and with time-varying delay. $a_{ij}\left(x_j(t)\right)$ and $b_{ij}\left(x_j\left(t-\tau_j(t)\right)\right)$ are memristive connective weights, which denote the neuron interconnection matrix and the delayed neuron interconnection matrix, respectively. W_{ij} and M_{ij} denote the memductances of memristors R_{ij} and F_{ij} respectively. And R_{ij} represents the memristor between the feedback function $f_i\left(x_i(t)\right)$ and $x_i(t)$, F_{ij} represents the memristor between the feedback function $g_i\left(x_i\left(t-\tau_i(t)\right)\right)$ and $x_i(t)$. I_i represents the external input. According to the feature of memristor, we denote

$$a_{ij}\left(x_j(t)\right) = \begin{cases} \hat{a}_{ij}, & x_j(t) \leq 0, \\ \check{a}_{ij}, & x_j(t) > 0, \end{cases} \quad b_{ij}\left(x_j\left(t-\tau_j(t)\right)\right) = \begin{cases} \hat{b}_{ij}, & x_j\left(t-\tau_j(t)\right) \leq 0. \\ \check{b}_{ij}, & x_j\left(t-\tau_j(t)\right) > 0. \end{cases} \tag{5}$$

C. Assumptions, Definitions and Lemmas

In the rest of paper, we first make following assumption for system (4) .

Assumption1: For $j \in N, \forall s_1, s_2 \in R$, the neuron activation functions f_j, g_j bounded, $f_j(0) = g_j(0) = 0$ and satisfy

$$0 \le \frac{f_j(s_1)-f_j(s_2)}{s_1-s_2} \le \sigma_j,$$
$$0 \le \frac{g_j(s_1)-g_j(s_2)}{s_1-s_2} \le \rho_j, \tag{6}$$

where $s_1 \neq s_2$ and σ_j, ρ_j are nonnegative constants.

We consider system (4) as drive system and corresponding response system is given as follows:3

$$D^q y_i(t) = -c_i y_i(t) + \sum_{j=1}^{n} a_{ij}\left(y_j(t)\right) f_j\left(y_j(t)\right) + \sum_{j=1}^{n} b_{ij}\left(y_j\left(t-\tau_j(t)\right)\right) g_j\left(y_j\left(t-\tau_j(t)\right)\right) + I_i + u_i,$$
$$t \ge 0,\ i \in N, \tag{7}$$

Where

$$a_{ij}\left(y_j(t)\right) = \begin{cases} \hat{a}_{ij}, & y_j(t) \le 0, \\ \check{a}_{ij}, & y_j(t) > 0, \end{cases} \quad b_{ij}\left(y_j\left(t-\tau_j(t)\right)\right) = \begin{cases} \hat{b}_{ij}, & y_j\left(t-\tau_j(t)\right) \le 0, \\ \check{b}_{ij}, & y_j\left(t-\tau_j(t)\right) > 0, \end{cases} \tag{8}$$

and $u_i(t)$ is a liner error feedback control function which defined by $u_i(t) = \omega_i\left(y_i(t) - x_i(t)\right)$, where $\omega_i,\ i \in N$ are constants, which denotes the control gain. Next, we define the synchronization error $e(t)$ as $e(t) = \left(e_1(t), e_2(t), ..., e_n(t)\right)^T$, where $e_i(t) = y_i(t) - x_i(t)$. According to the system (4) and system (7), the synchronization error system can be described as follows:

$$D^q e_i(t) = -c_i e_i(t) + \left[\sum_{j=1}^{n} a_{ij}\left(y_j(t)\right) f_j\left(y_j(t)\right) - \sum_{j=1}^{n} a_{ij}\left(x_j(t)\right) f_j\left(x_j(t)\right)\right]$$
$$+\left[\sum_{j=1}^{n} b_{ij}\left(y_j\left(t-\tau_j(t)\right)\right) g_j\left(y_j\left(t-\tau_j(t)\right)\right) - \sum_{j=1}^{n} b_{ij}\left(x_j\left(t-\tau_j(t)\right)\right) g_j\left(x_j\left(t-\tau_j(t)\right)\right)\right]$$
$$+u_i(t),\ t \ge 0, i \in N \tag{9}$$

where $a_{ij}\left(y_j(t)\right), b_{ij}\left(y_j\left(t-\tau_j(t)\right)\right), a_{ij}\left(x_j(t)\right), b_{ij}\left(x_j\left(t-\tau_j(t)\right)\right)$ are the same as those defined above, $u_i(t) = \omega_i\left(y_i(t) - x_i(t)\right) = \omega_i e_i(t)$, where $\omega_i,\ i \in N$ are constants, which denotes the control gain.

According to the theories of differential inclusions and set valued maps [40], if $x_i(t)$ and $y_i(t)$ are solutions of (4) and (7) respectively, system (4) and system (7) can be written as follow:

$$D^{q}x_i(t)\in -c_i x_i(t)+\sum_{j=1}^{n} co\left[a_{ij}\left(x_j(t)\right)\right]f_j\left(x_j(t)\right)$$
$$+\sum_{j=1}^{n} co\left[b_{ij}\left(x_j\left(t-\tau_j(t)\right)\right)\right]g_j\left(x_j\left(t-\tau_j(t)\right)\right)+I_i,\ t\geq 0, i\in N \tag{10}$$

And

$$D^{q}y_i(t)\in -c_i y_i(t)+\sum_{j=1}^{n} co\left[a_{ij}\left(y_j(t)\right)\right]f_j\left(y_j(t)\right)$$
$$+\sum_{j=1}^{n} co\left[b_{ij}\left(y_j\left(t-\tau_j(t)\right)\right)\right]g_j\left(y_j\left(t-\tau_j(t)\right)\right)+I_i+u_i,\ t\geq 0, i\in N, \tag{11}$$

Where

$$co\left[a_{ij}\left(x_j(t)\right)\right]=\begin{cases}\hat{a}_{ij}, x_j(t)<0,\\ co\left\{\hat{a}_{ij},\check{a}_{ij}\right\},\ x_j(t)=0,\\ \check{a}_{ij}, x_j(t)>0,\end{cases}\quad co\left[a_{ij}\left(y_j(t)\right)\right]=\begin{cases}\hat{a}_{ij}, y_j(t)<0\\ co\left\{\hat{a}_{ij},\check{a}_{ij}\right\},\ y_j(t)=0\\ \check{a}_{ij}, y_j(t)>0\end{cases} \tag{12}$$

And

$$co\left[b_{ij}\left(x_j\left(t-\tau(j)\right)\right)\right]=\begin{cases}\hat{b}_{ij}, x_j\left(t-\tau(j)\right)<0,\\ co\left\{\hat{b}_{ij},\check{b}_{ij}\right\}, x_j\left(t-\tau(j)\right)=0,\\ \check{b}_{ij}, x_j\left(t-\tau(j)\right)>0,\end{cases}$$
$$co\left[b_{ij}\left(y_j\left(t-\tau(j)\right)\right)\right]=\begin{cases}\hat{b}_{ij}, y_j\left(t-\tau(j)\right)<0,\\ co\left\{\hat{b}_{ij},\check{b}_{ij}\right\}, y_j\left(t-\tau(j)\right)=0,\\ \check{b}_{ij}, y_j\left(t-\tau(j)\right)>0,\end{cases} \tag{13}$$

where $co\{u,v\}$ denotes the closure of convex hull generated by real numbers u and v or real matrices u and v.

Then the synchronization error system can be described as follows:

$$D^{q}e_i(t) \in -c_i e_i(t) + \left\{\sum_{j=1}^{n} co\left[a_{ij}\left(y_j(t)\right)\right] f_j\left(y_j(t)\right) - \sum_{j=1}^{n} co\left[a_{ij}\left(x_j(t)\right)\right] f_j\left(x_j(t)\right)\right\}$$
$$+\left\{\sum_{j=1}^{n} co\left[b_{ij}\left(y_j\left(t-\tau_j(t)\right)\right)\right] g_j\left(y_j\left(t-\tau_j(t)\right)\right)\right.$$
$$\left.-\sum_{j=1}^{n} co\left[b_{ij}\left(x_j\left(t-\tau_j(t)\right)\right)\right] g_j\left(x_j\left(t-\tau_j(t)\right)\right)\right\} + \omega_i e_i,\ t \geq 0, i \in N. \tag{14}$$

Definition2 [8] For $\forall t \geq 0,$ the exponential synchronization of system (4) and system (7) can be transformed to the exponential stability of the error system (9) (error approaches to zero). The error system (9) is said to be exponentially stable, if there exist constant $Q_i > 0, \quad P_i > 0,$ such that the solution $e(t) = \left(e_1(t), e_2(t), ..., e_n(t)\right)^T$ of error system (9) with initial condition $e(s) = \phi(s) \in \left(\left[t_0 - \tau, t_0\right], R^n\right)$ satisfies

$$\left|e_i(t)\right| \leq Q_i \max_{1 \leq i \leq n}\left\{\sup_{t_0 - \tau \leq s \leq t_0}\left|\phi_i(s)\right|\right\} \exp\left\{-P_i\left(t - t_0\right)\right\},\ t \geq t_0 > 0,$$

$i = 1, 2, ..., n$, where P_i is called the estimated rate of exponential convergence.

Lemma1 [14] Under the assumption1, the following estimation can be obtained:

(i) $co\left[a_{ij}\left(y_j(t)\right)\right] f_j\left(y_j(t)\right) - co\left[a_{ij}\left(x_j(t)\right)\right] f_j\left(x_j(t)\right) \leq A_{ij} F_j\left(e_j(t)\right)$,

(ii) $co\left[b_{ij}\left(y_j\left(t-\tau_j(t)\right)\right)\right] g_j\left(y_j\left(t-\tau_j(t)\right)\right) - co\left[b_{ij}\left(x_j\left(t-\tau_j(t)\right)\right)\right] g_j\left(x_j\left(t-\tau_j(t)\right)\right)$

$\leq B_{ij} G_j\left(e_j\left(t-\tau_j(t)\right)\right)$,

where $A_{ij} = \max\left\{\left|\hat{a}_{ij}\right|, \left|\check{a}_{ij}\right|\right\}, B_{ij} = \max\left\{\left|\hat{b}_{ij}\right|, \left|\check{b}_{ij}\right|\right\}, i, j \in N,$

$F_j\left(e_j(t)\right) = f_j\left(y_j(t)\right) - f_j\left(x_j(t)\right), G_j\left(e_j\left(t-\tau_j(t)\right)\right) = g_j\left(y_j\left(t-\tau_j(t)\right)\right) - g_j\left(x_j\left(t-\tau_j(t)\right)\right), j \in N.$

Proof: If $y_i(t) = 0, x_i(t) = 0, i \in N$ we can easily have part(i) hold. From (9) and(10), we can get

(1) For $y_i(t) < 0, x_i(t) < 0$, then

$co\left[a_{ij}\left(y_j(t)\right)\right] f_j\left(y_j(t)\right) - co\left[a_{ij}\left(x_j(t)\right)\right] f_j\left(x_j(t)\right) = \hat{a}_{ij} f_j\left(y_j(t)\right) - \hat{a}_{ij} f_j\left(x_j(t)\right)$

$= \hat{a}_{ij} F_j\left(e_j(t)\right) \leq A_{ij} F_j\left(e_j(t)\right).$

(2) For $y_i(t) > 0, x_i(t) > 0$, then

$co\left[a_{ij}\left(y_j(t)\right)\right] f_j\left(y_j(t)\right) - co\left[a_{ij}\left(x_j(t)\right)\right] f_j\left(x_j(t)\right) = \check{a}_{ij} f_j\left(y_j(t)\right) - \check{a}_{ij} f_j\left(x_j(t)\right)$

$$= \breve{a}_{ij} F_j\left(e_j(t)\right) \le A_{ij} F_j\left(e_j(t)\right).$$

(3) For $x_i(t) < 0 < y_i(t)$ or $y_i(t) < 0 < x_i(t)$, then

$$co\left[a_{ij}\left(y_j(t)\right)\right] f_j\left(y_j(t)\right) - co\left[a_{ij}\left(x_j(t)\right)\right] f_j\left(x_j(t)\right)$$
$$= \breve{a}_{ij}\left(f_j\left(y_j(t)\right) - f(0)\right) + \hat{a}_{ij}\left(f(0) - f_j\left(x_j(t)\right)\right)$$
$$\le A_{ij}\left(f_j\left(y_j(t)\right) - f(0)\right) + A_{ij}\left(f(0) - f_j\left(x_j(t)\right)\right) = A_{ij}\left(f_j\left(y_j(t)\right) - f_j\left(x_j(t)\right)\right) = A_{ij} F_j\left(e_j(t)\right).$$

Then complete the proof of part (i). In the similar way, part(ii) can be easily hold.

III. MAIN RESULTS

We present the exponential stability results for the synchronization error system of FMNN，when the error system (9) is exponentially stable, the system (4) and system (7) will achieve the exponential synchronization.

Theorem1 If there exist positive constant $\varepsilon, \eta_1, \eta_2, ..., \eta_n$ such that for any $t \ge t_0 > 0, i \in \{1, 2, ..., n\}$

$$(-c_i + \omega_i + \varepsilon)\eta_i + \sum_{j=1}^{n} A_{ij}\sigma_j\eta_j + \sum_{j=1}^{n} B_{ij}\rho_j\eta_j \exp\{\varepsilon\tau_j(t)\} < 0, \tag{15}$$

then the error system (9) is globally exponentially stable.

Proof: Consider $W_i(t) = |e_i(t)|/\eta_i, i = 1, 2, ..., n$, according to the error system (9) or (14) and lemma1, we can get the following inequality

$$D^q e_i(t) \le -c_i e_i(t) + \sum_{j=1}^{n} A_{ij} F_j\left(e_j(t)\right) + \sum_{j=1}^{n} B_{ij} G_j\left(e_j\left(t - \tau_j(t)\right)\right) + \omega_i e_i(t)$$
$$= (-c_i + \omega_i) e_i(t) + \sum_{j=1}^{n} A_{ij} F_j\left(e_j(t)\right) + \sum_{j=1}^{n} B_{ij} G_j\left(e_j\left(t - \tau_j(t)\right)\right). \tag{16}$$

Evaluating the fractional order derivative of $W_i(t)$ along the trajectory of error system, then

$$D^q W_i(t) \le (-c_i + \omega_i)|e_i(t)|/\eta_i + 1/\eta_i\left[\sum_{j=1}^{n} A_{ij}\sigma_j |e_j(t)| + \sum_{j=1}^{n} B_{ij}\rho_j \left|e_j\left(t - \tau_j(t)\right)\right|\right]$$
$$= (-c_i + \omega_i) W_i(t) + 1/\eta_i\left[\sum_{j=1}^{n} A_{ij}\sigma_j\eta_j W_j(t) + \sum_{j=1}^{n} B_{ij}\rho_j\eta_j W_j\left(t - \tau_j(t)\right)\right]. \tag{17}$$

Define $\tilde{W}_i(t) = W_i(t) - \bar{W}(t_0)\exp\{-\varepsilon(t - t_0)\}, t \ge t_0 > 0, i = 1, 2, ..., n$, where

$$\bar{W}(t_0)=\max_{1\le i\le n}\left\{\sup_{t_0-\tau\le s\le t_0}\left|e_i(s)\right|/\eta_i\right\}.$$

We will prove that $\tilde{W}_i(t)\le 0, i=1,2,...,n$, for any $t\ge t_0>0$. Otherwise, since $\tilde{W}_i(t)\le 0,\ i=1,2,...,n$ for $t\in[t_0-\tau,\ t_0]$, there must exist $t_1\ge t_0$ and some ς such that $D^q\tilde{W}_\varsigma(t_1)\ge 0$ and $\tilde{W}_\varsigma(t_1)=0$. Then

$$\begin{aligned}
&D^q\tilde{W}_\varsigma(t_1)\le(-c_\varsigma+\omega_\varsigma)W_\varsigma(t_1)+1/\eta_\varsigma\left[\sum_{j=1}^n A_{\varsigma j}\sigma_j\eta_j W_j(t_1)+\sum_{j=1}^n B_{\varsigma j}\rho_j\eta_j W_j\left(t-\tau_j(t_1)\right)\right]\\
&+\varepsilon\bar{W}(t_0)\exp\{-\varepsilon(t_1-t_0)\}\\
&=(-c_\varsigma+\omega_\varsigma)\bar{W}(t_0)\exp\{-\varepsilon(t_1-t_0)\}+1/\eta_\varsigma\left[\sum_{j=1}^n A_{\varsigma j}\sigma_j\eta_j\bar{W}(t_0)\exp\{-\varepsilon(t_1-t_0)\}\right.\\
&\left.+\sum_{j=1}^n B_{\varsigma j}\rho_j\eta_j\bar{W}(t_0)\exp\left\{-\varepsilon\left(t_1-\tau_j(t_1)-t_0\right)\right\}\right]+\varepsilon\bar{W}(t_0)\exp\{-\varepsilon(t_1-t_0)\}\\
&=(-c_\varsigma+\omega_\varsigma+\varepsilon)\bar{W}(t_0)\exp\{-\varepsilon(t_1-t_0)\}+1/\eta_\varsigma\left[\sum_{j=1}^n A_{\varsigma j}\sigma_j\eta_j\bar{W}(t_0)\exp\{-\varepsilon(t_1-t_0)\}\right.\\
&\left.+\sum_{j=1}^n B_{\varsigma j}\rho_j\eta_j\bar{W}(t_0)\exp\left\{-\varepsilon\left(t_1-\tau_j(t_1)-t_0\right)\right\}\right]\\
&=(-c_\varsigma+\omega_\varsigma+\varepsilon)\bar{W}(t_0)\exp\{-\varepsilon(t_1-t_0)\}\\
&+1/\eta_\varsigma\left[\sum_{j=1}^n A_{\varsigma j}\sigma_j\eta_j\bar{W}(t_0)+\sum_{j=1}^n B_{\varsigma j}\rho_j\eta_j\bar{W}(t_0)\exp\{\varepsilon\tau_j(t_1)\}\right]\exp\{-\varepsilon(t_1-t_0)\}.
\end{aligned}\tag{18}$$

Moreover, from inequality(15), we have

$$(-c_i+\omega_i+\varepsilon)+1/\eta_i\left[\sum_{j=1}^n A_{ij}\sigma_j\eta_j+\sum_{j=1}^n B_{ij}\rho_j\eta_j\exp\{\varepsilon\tau_j(t)\}\right]<0, t\ge t_0>0, i=1,2,...,n,$$

Therefore

$$\begin{aligned}
&(-c_i+\omega_i+\varepsilon)\bar{W}(t_0)\exp\{-\varepsilon(t_1-t_0)\}\\
&+1/\eta_i\left[\sum_{j-1}^n A_{ij}\sigma_j\eta_j+\sum_{j-1}^n B_{ij}\rho_j\eta_j\exp\{\varepsilon\tau_j(t)\}\right]\bar{W}(t_0)\exp\{-\varepsilon(t_1-t_0)\}<0,\\
&t\ge t_0>0, i=1,2,...,n,
\end{aligned}\tag{19}$$

so it is easy to find that $D^q\tilde{W}_\varsigma(t_1)<0$, which contradicts $D^q\tilde{W}_\varsigma(t_1)\ge 0$. That shows $\tilde{W}_i(t)=W_i(t)-\bar{W}(t_0)\exp\{-\varepsilon(t-t_0)\}\le 0,\ t\ge t_0>0, i=1,2,...,n$. Thus

$$\left|e_i(t)\right|/\eta_i\le\max_{1\le i\le n}\left\{\sup_{t_0-\tau\le s\le t_0}\left|e_i(s)\right|/\eta_i\right\}\exp\{-\varepsilon(t-t_0)\},\ t\ge t_0>0, i=1,2,...,n.$$

It shows

$$\left|e_i(t)\right| \le \eta_i \max_{1\le i\le n}\left\{\sup_{t_0-\tau\le s\le t_0}\left|e_i(s)\right|/\eta_i\right\}\exp\left\{-\varepsilon\left(t-t_0\right)\right\},\ t\ge t_0>0, i=1,2,...,n. \tag{20}$$

This completes the proof.

IV. NUMERICAL RESULTS

In this section, we will give two numerical examples to demonstrate our analysis on exponential synchronization of FMNN.

Example1 Consider two-dimension fractional-order memristor-based neural networks

$$\begin{cases} D^q x_1(t) = -c_1 x_1(t) + a_{11}\left(x_1(t)\right) f_1\left(x_1(t)\right) + a_{12}\left(x_2(t)\right) f_2\left(x_2(t)\right) \\ +b_{11}\left(x_1\left(t-\tau_1(t)\right)\right) g_1\left(x_1\left(t-\tau_1(t)\right)\right) + b_{12}\left(x_2\left(t-\tau_2(t)\right)\right) g_2\left(x_2\left(t-\tau_2(t)\right)\right) + I_1 \\ D^q x_2(t) = -c_2 x_2(t) + a_{21}\left(x_1(t)\right) f_1\left(x_1(t)\right) + a_{22}\left(x_2(t)\right) f_2\left(x_2(t)\right) \\ +b_{21}\left(x_1\left(t-\tau_1(t)\right)\right) g_1\left(x_1\left(t-\tau_1(t)\right)\right) + b_{22}\left(x_2\left(t-\tau_2(t)\right)\right) g_2\left(x_2\left(t-\tau_2(t)\right)\right) + I_2 \end{cases} \tag{21}$$

where $c_1 = c_2 = 1,\ a_{11}\left(x_1(t)\right) = 1,\ a_{22}\left(x_2(t)\right) = 1.8,$

$$a_{12}\left(x_2(t)\right) = \begin{cases} 12, & x_2(t)\le 0, \\ 14, & x_2(t)>0, \end{cases} \quad a_{21}\left(x_1(t)\right) = \begin{cases} 0.1, & x_1(t)\le 0, \\ 0.05, & x_1(t)>0, \end{cases}$$

$$b_{11}\left(x_1\left(t-\tau_1(t)\right)\right) = \begin{cases} -1.2, & x_1\left(t-\tau_1(t)\right)\le 0, \\ -1.5, & x_1\left(t-\tau_1(t)\right)>0, \end{cases} \quad b_{12}\left(x_2\left(t-\tau_2(t)\right)\right) = \begin{cases} 0.8, & x_2\left(t-\tau_2(t)\right)\le 0, \\ 1.0, & x_2\left(t-\tau_2(t)\right)>0, \end{cases}$$

$$b_{21}\left(x_1\left(t-\tau_1(t)\right)\right) = \begin{cases} 0.05, & x_1\left(t-\tau_1(t)\right)\le 0, \\ 0.1, & x_1\left(t-\tau_1(t)\right)>0, \end{cases} \quad b_{22}\left(x_2\left(t-\tau_2(t)\right)\right) = \begin{cases} -1.6, & x_2\left(t-\tau_2(t)\right)\le 0, \\ -1.4, & x_2\left(t-\tau_2(t)\right)>0, \end{cases}$$

where $\tau_j(t) = e^t/1+e^t,\ I = (I_1, I_2)^T = (0,0)^T,$ $q = 0.92$ and take the activation function as $f_i(x_i) = \sin(x_i),$ $g_i(x_i) = 0.5\left(\left|x_i+1\right| - \left|x_i-1\right|\right),\ i,\ j = 1,2.$ The model (21) has chaotic attractors with initial values $x(0) = (0.45, 0.65)^T$ which can be seen in Figure1.

We consider system (21) as the drive system and corresponding response system is defined as Eq.(7). And for the controller $u_i(t) = \omega_i\left(y_i(t) - x_i(t)\right),$ the parameter ω_i is chosen as $\omega_1 = -9.5,\ \omega_2 = -10.5.$ From Theorem1, when we take $\varepsilon = 0.7, \tau_j(t) = 1,$ $\eta_1 = \eta_2 = \rho_1 = \rho_2 = \sigma_1 = \sigma_2 = 0.1,$ we can easily know

$(-c_i+\omega_i+\varepsilon)\eta_i+\sum_{j=1}^{n}A_{ij}\sigma_j\eta_j+\sum_{j=1}^{n}B_{ij}\rho_j\eta_j\exp\{\varepsilon\tau_j(t)\}<0 \quad \omega_1=-9.5,\omega_2=-10.5,$ we can get

is true when $\omega_1<-1.703,\omega_2<-0.232$. So when

$$(-c_1+\omega_1+\varepsilon)\eta_1+A_{11}\sigma_1\eta_1+A_{12}\sigma_2\eta_2+B_{11}\rho_1\eta_1\exp\{\varepsilon\tau_1(t)\}+B_{12}\rho_2\eta_2\exp\{\varepsilon\tau_2(t)\}=-0.798<0,$$
$$(-c_2+\omega_2+\varepsilon)\eta_2+A_{21}\sigma_1\eta_1+A_{22}\sigma_2\eta_2+B_{21}\rho_1\eta_1\exp\{\varepsilon\tau_1(t)\}+B_{22}\rho_2\eta_2\exp\{\varepsilon\tau_2(t)\}=-1.027<0.$$

It satisfies the condition of Theorem 1, then the exponential synchronization of drive-response system is achieved.

When the response system with this controller, we get state trajectories of variable $x_1(t),y_1(t)$ and $x_2(t),y_2(t)$ are depicted in Figure2a and 2b. Moreover, Figure3a and 3b depict the synchronization error curves $e_1(t),e_2(t)$ between the drive system and response system. These numerical simulations show the state trajectories of variable $x_1(t),y_1(t)$ and $x_2(t),y_2(t)$ are synchronous and synchronization error $e_1(t),e_2(t)$ are converge to zero. These prove the correctness of the Theorem1.

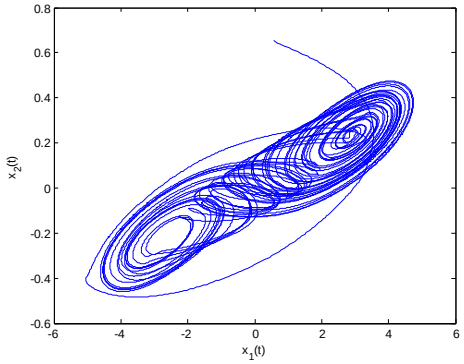


Figure 1. The chaotic attractors of fractional-order memristor-based neural networks(18)

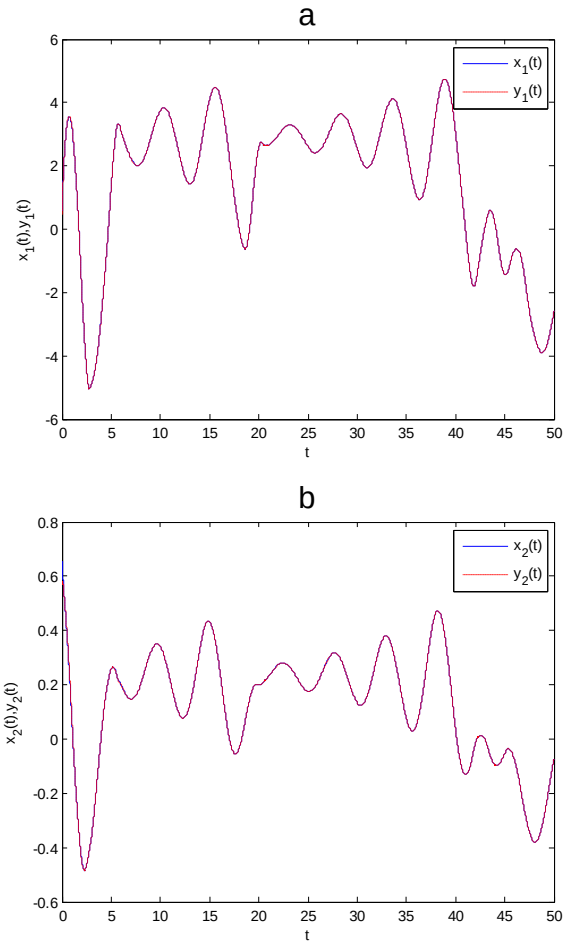


Figure 2. Exponential synchronization of state variable with cntroller $(a:x_1(t),y_1(t),b:x_2(t),y_2(t))$

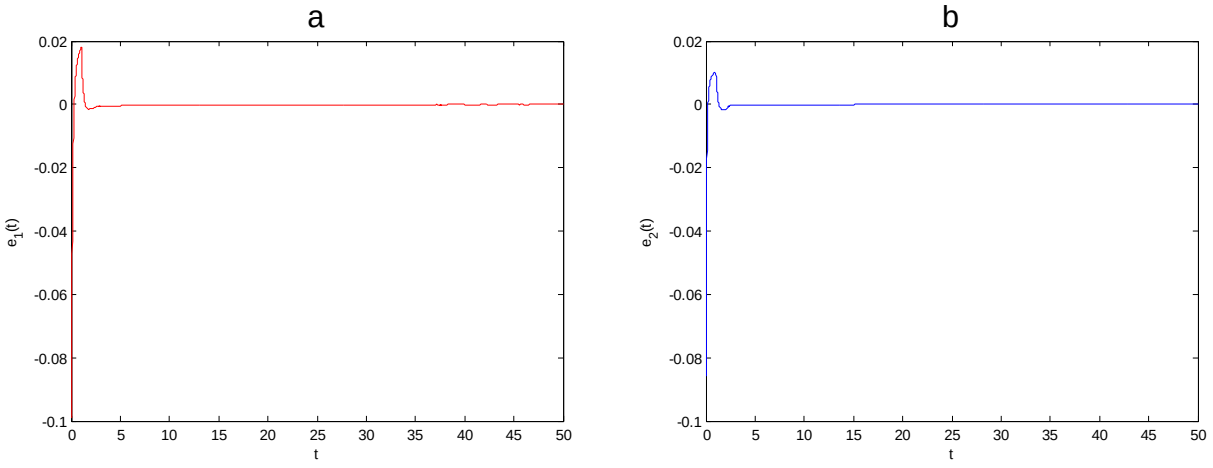


Figure 3. Synchronization error between the drive and response system $\left(a:e_1(t),b:e_2(t)\right)$

Example2 Consider three-dimension fractional-order memristor-based neural networks

$$\begin{cases} D^q x_1(t) = -c_1 x_1(t) + a_{11}\left(x_1(t)\right) f_1\left(x_1(t)\right) + a_{12}\left(x_2(t)\right) f_2\left(x_2(t)\right) + a_{13}\left(x_3(t)\right) f_3\left(x_3(t)\right) \\ +b_{11}\left(x_1\left(t-\tau_1(t)\right)\right) g_1\left(x_1\left(t-\tau_1(t)\right)\right) + b_{12}\left(x_2\left(t-\tau_2(t)\right)\right) g_2\left(x_2\left(t-\tau_2(t)\right)\right) \\ +b_{13}\left(x_3\left(t-\tau_3(t)\right)\right) g_3\left(x_3\left(t-\tau_3(t)\right)\right) + I_1 \\ D^q x_2(t) = -c_2 x_2(t) + a_{21}\left(x_1(t)\right) f_1\left(x_1(t)\right) + a_{22}\left(x_2(t)\right) f_2\left(x_2(t)\right) + a_{23}\left(x_3(t)\right) f_3\left(x_3(t)\right) \\ +b_{21}\left(x_1\left(t-\tau_1(t)\right)\right) g_1\left(x_1\left(t-\tau_1(t)\right)\right) + b_{22}\left(x_2\left(t-\tau_2(t)\right)\right) g_2\left(x_2\left(t-\tau_2(t)\right)\right) \\ +b_{23}\left(x_3\left(t-\tau_3(t)\right)\right) g_3\left(x_3\left(t-\tau_3(t)\right)\right) + I_2 \\ D^q x_3(t) = -c_3 x_3(t) + a_{31}\left(x_1(t)\right) f_1\left(x_1(t)\right) + a_{32}\left(x_2(t)\right) f_2\left(x_2(t)\right) + a_{33}\left(x_3(t)\right) f_3\left(x_3(t)\right) \\ +b_{31}\left(x_1\left(t-\tau_1(t)\right)\right) g_1\left(x_1\left(t-\tau_1(t)\right)\right) + b_{32}\left(x_2\left(t-\tau_2(t)\right)\right) g_2\left(x_2\left(t-\tau_2(t)\right)\right) \\ +b_{33}\left(x_3\left(t-\tau_3(t)\right)\right) g_3\left(x_3\left(t-\tau_3(t)\right)\right) + I_3 \end{cases} \tag{22}$$

where $c_1 = c_2 = c_3 = 1,$

$$a_{11}\left(x_1(t)\right) = \begin{cases} 1, & x_1(t) \le 0, \\ -1, & x_1(t) > 0, \end{cases} \quad a_{21}\left(x_1(t)\right) = \begin{cases} 1, & x_1(t) \le 0, \\ -1, & x_1(t) > 0, \end{cases} \quad a_{31}\left(x_1(t)\right) = \begin{cases} 1, & x_1(t) \le 0, \\ -1, & x_1(t) > 0, \end{cases}$$

$$a_{12}\left(x_2(t)\right) = \begin{cases} -1, & x_2(t) \le 0, \\ 1, & x_2(t) > 0, \end{cases} \quad a_{22}\left(x_2(t)\right) = \begin{cases} 1, & x_2(t) \le 0, \\ -1, & x_2(t) > 0, \end{cases} \quad a_{32}\left(x_2(t)\right) = \begin{cases} -1, & x_2(t) \le 0, \\ 1, & x_2(t) > 0, \end{cases}$$

$$a_{13}\left(x_3(t)\right) = \begin{cases} -1, & x_3(t) \le 0, \\ 1, & x_3(t) > 0, \end{cases} \quad a_{23}\left(x_3(t)\right) = \begin{cases} 1, & x_3(t) \le 0, \\ -1, & x_3(t) > 0, \end{cases} \quad a_{33}\left(x_3(t)\right) = \begin{cases} 1, & x_3(t) \le 0, \\ -1, & x_3(t) > 0, \end{cases}$$

$$b_{11}\left(x_1\left(t-\tau_1(t)\right)\right)=\begin{cases}1, & x_1\left(t-\tau_1(t)\right)\le 0,\\ -1, & x_1\left(t-\tau_1(t)\right)>0,\end{cases}$$

$$b_{12}\left(x_2\left(t-\tau_2(t)\right)\right)=\begin{cases}-1, & x_2\left(t-\tau_2(t)\right)\le 0,\\ 1, & x_2\left(t-\tau_2(t)\right)>0,\end{cases}$$

$$b_{21}\left(x_1\left(t-\tau_1(t)\right)\right)=\begin{cases}1, & x_1\left(t-\tau_1(t)\right)\le 0,\\ -1, & x_1\left(t-\tau_1(t)\right)>0,\end{cases}$$

$$b_{22}\left(x_2\left(t-\tau_2(t)\right)\right)=\begin{cases}1, & x_2\left(t-\tau_2(t)\right)\le 0,\\ -1, & x_2\left(t-\tau_2(t)\right)>0,\end{cases}$$

$$b_{31}\left(x_2\left(t-\tau_2(t)\right)\right)=\begin{cases}1, & x_2\left(t-\tau_2(t)\right)\le 0,\\ -1, & x_2\left(t-\tau_2(t)\right)>0,\end{cases}$$

$$b_{32}\left(x_2\left(t-\tau_2(t)\right)\right)=\begin{cases}-1, & x_2\left(t-\tau_2(t)\right)\le 0,\\ 1, & x_2\left(t-\tau_2(t)\right)>0,\end{cases}$$

$$b_{13}\left(x_3\left(t-\tau_3(t)\right)\right)=\begin{cases}-1, & x_3\left(t-\tau_3(t)\right)\le 0,\\ 1, & x_3\left(t-\tau_3(t)\right)>0,\end{cases}$$

$$b_{23}\left(x_3\left(t-\tau_3(t)\right)\right)=\begin{cases}1, & x_3\left(t-\tau_3(t)\right)\le 0,\\ -1, & x_3\left(t-\tau_3(t)\right)>0,\end{cases}$$

$$b_{33}\left(x_3\left(t-\tau_3(t)\right)\right)=\begin{cases}1, & x_3\left(t-\tau_3(t)\right)\le 0,\\ -1, & x_3\left(t-\tau_3(t)\right)>0.\end{cases}$$

And $\tau_j(t)=e^t/1+e^t, I=(I_1,I_2,I_3)^T=(0,0,0)^T$ $q=0.92$ and take the activation function as $f_i(x_i)=g_i(x_i)=\tanh(x_i), i=1,2,3$ .We consider system(22) as the drive system and the corresponding response system is defined in Eq.(7). And for the controller $u_i(t)=\omega_i\left(y_i(t)-x_i(t)\right),$ ω_i is chosen as $\omega_1=-9.5,\omega_2=-10.5,\omega_3=-11$. From Theorem1, we take $\varepsilon=0.7,\tau_j(t)=1$ and choose $\eta_1=\eta_2=0.1$ $\rho_1=\rho_2=\sigma_1=\sigma_2=0.1.$ According to $A_{ij}=\max\left\{\left|\hat{a}_{ij}\right|,\left|\check{a}_{ij}\right|\right\}, B_{ij}=\max\left\{\left|\hat{b}_{ij}\right|,\left|\check{b}_{ij}\right|\right\}$ $i,j=1,2,3$ $A_{ij}=B_{ij}=1,$ we can easily know $(-c_i+\omega_i+\varepsilon)\eta_i+\sum_{j=1}^{n}A_{ij}\sigma_j\eta_j+\sum_{j=1}^{n}B_{ij}\rho_j\eta_j\exp\left\{\varepsilon\tau_j(t)\right\}<0$ is true when $\omega_i<-0.604.$ So when $\omega_1=-9.5,\ \omega_2=-10.5,\omega_3=-11$ we can get

$$(-c_1+\omega_1+\varepsilon)\eta_1+A_{11}\sigma_1\eta_1+A_{12}\sigma_2\eta_2+A_{13}\sigma_3\eta_3+\left(B_{11}\rho_1\eta_1+B_{12}\rho_2\eta_2+B_{13}\rho_3\eta_3\right)\exp\left\{\varepsilon\tau_j(t)\right\}=-0.89<0,$$

$$(-c_2+\omega_2+\varepsilon)\eta_1+A_{21}\sigma_1\eta_1+A_{22}\sigma_2\eta_2+A_{23}\sigma_3\eta_3+\left(B_{21}\rho_1\eta_1+B_{22}\rho_2\eta_2+B_{23}\rho_3\eta_3\right)\exp\left\{\varepsilon\tau_j(t)\right\}=-0.99<0,$$

$$(-c_3+\omega_3+\varepsilon)\eta_1+A_{31}\sigma_1\eta_1+A_{32}\sigma_2\eta_2+A_{33}\sigma_3\eta_3+\left(B_{31}\rho_1\eta_1+B_{32}\rho_2\eta_2+B_{33}\rho_3\eta_3\right)\exp\left\{\varepsilon\tau_j(t)\right\}=-1.04<0.$$

It suggests the condition of Theorem 1 is satisfied, then drive-response system achieves the synchronization.

When the response system with this controller, we get state trajectories of variable $x_1(t),y_1(t)$ and $x_2(t),y_2(t)$

and $x_3(t), y_3(t)$ are depicted in Figure 4a,4b,4c. Moreover, Figure 5a,5b,5c depict the synchronization error curves $e_1(t), e_2(t), e_3(t)$ between the drive system and response system. It's easy to see that the state trajectories of variable $x_1(t), y_1(t)$, $x_2(t), y_2(t)$ and $x_3(t), y_3(t)$ are synchronous and synchronization error $e_1(t), e_2(t), e_3(t)$ are converge to zero. So the Theorem1 is proved to be correct.

In addition, we choose $\omega_1 = -9.5$, $\omega_2 = -10.5, \omega_3 = -11$, according to the Theorem1, it needs the following inequalities to hold:

$$\begin{cases} \tau < \frac{1}{\varepsilon} \ln\left(\frac{117}{3} - \frac{10}{3}\varepsilon\right) \\ \tau < \frac{1}{\varepsilon} \ln\left(\frac{112}{3} - \frac{10}{3}\varepsilon\right) \\ \tau < \frac{1}{\varepsilon} \ln\left(\frac{102}{3} - \frac{10}{3}\varepsilon\right) \end{cases}$$

So, we just need $\tau < \frac{1}{\varepsilon} \ln\left(\frac{102}{3} - \frac{10}{3}\varepsilon\right)$ holds. We have the exponential convergence rate $0 < \varepsilon < 1$, figure 6 depicts the relation of time-varying delay τ and exponential convergence rate ε.

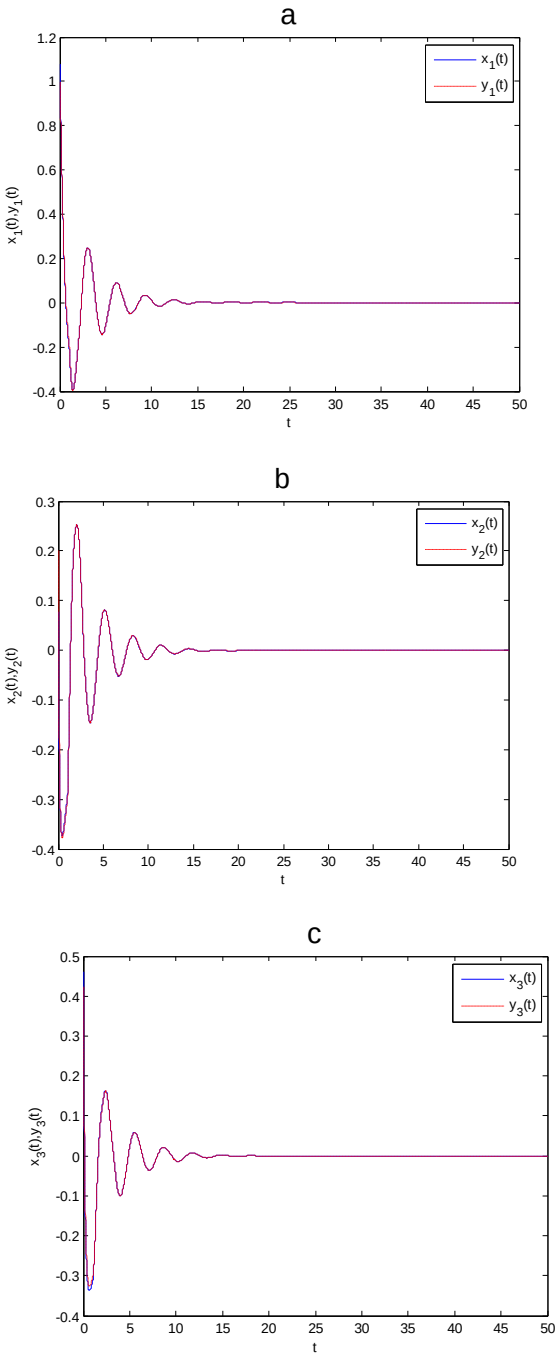


Figure 4. Synchronization of state variable with controller $(a: x_1(t), y_1(t), b: x_2(t), y_2(t), c: x_3(t), y_3(t))$

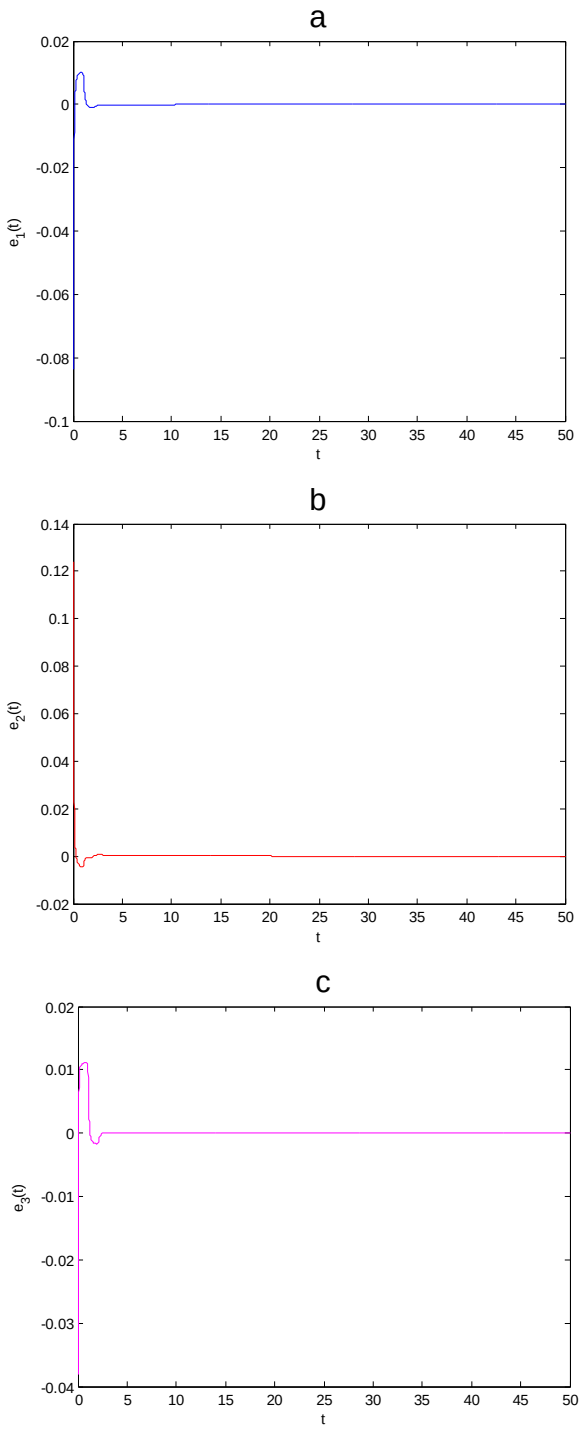


Figure 5. Synchronization error between the drive and response system $\left(a:e_1(t),b:e_2(t),c:e_3(t)\right)$

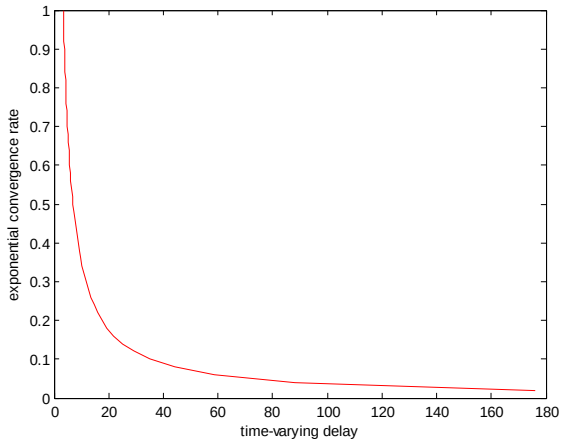


Figure 6. The relation of time-varying delay τ and exponential convergence rate ε .

V. CONCLUSION

This paper achieves the exponential synchronization of a class of FMNN with time-varying delays by using linear error feedback controller. Based on comparison principle, the new theorem is derived to guarantee the exponential synchronization between the drive system and response system. The methods proposed for synchronization is effective and it is easy to achieve than other complex control methods. Moreover, it can be extended to investigate other dynamical behaviors of fractional-order memristive neural networks, such as realizing the lag synchronization or anti-synchronizaton of this system based on the suitable controller. These issues will be the topic of future research. Finally, numerical examples are given to illustrate the effectiveness of the proposed theory.

REFERENCES

[1] L.O. Chua, "Memristor-the missing circuit element," IEEE Trans Circuit Theory, vol.18, pp.507–519, 1971.

[2] D.B.Strukov, G.S.Snider and D.R.Stewart, "The missing memristor found," Nature, vol.453, pp.80-83, 2008.

[3] L.O.Chua, "Resistance switching memories are memristors," Applied Physics A Materials Science and Processing, vol.102,pp.765-783, 2011.

[4] M.J.Sharifiy and Y.M.Banadaki, "General spice models for memristor and application to circuit simulation of memristor-based synapses and memory cells," J Circuit Syst Comp, vol.19, pp.407-424, 2010.

[5] Y.V.Pershin, V.M.Di, "Experimental demonstration of associative memory with memristive neural networks,"Neural Networks, vol.23, pp.881-886, 2010.

[6] M.Itoh and L.O.Chua, "Memristor cellular automata and memristor discrete-time cellular neural networks," Int J Bifur Chaos,vol.19, pp. 3605–3656, 2010.

[7] G.D.Zhang, Y.Shen and J.Sun, "Global exponential stability of a class of memristor-based recurrent neural networks with time-varying delays, "Neurocomputing, vol.97, pp.149–154, 2012.

[8] A.L.Wu and Z.G.Zeng, "New global exponential stability results for a memristive neural system with time-varying delays," Neurocomputing, vol.144, pp.553–559, 2014.

[9] A.L.Wu and Z.G.Zeng. "Exponential stabilization of memristive neural networks with time delays," IEEE Trans on Neural Netw and Learn Syst, vol.23, pp.1919–1929, 2012.

[10] A.L.Wu and Z.G.Zeng, "Lagrange stability of neural networks with memristive synapses and multiple delays," Inf Sci, vol.280, pp.135–151, 2014.

[11] G.D.Zhang, Y. Shen, Q.Yin and J.W.Sun, "Global exponential periodicity and stability of a class of memristor-based recurrent neural networks with multiple delays,"Inf Sci, vol.232, pp.386–396, 2013.

[12] J.J.Chen, Z.G.Zeng and P. Jiang, "On the periodic dynamics of memristor-based neural networks with time-varying delays," Inf Sci, vol.279, pp.358–373, 2014.

[13] L.Duan and L.H.Huang, "Existence and stability of periodic solution for mixed time-varying delayed neural networks with discontinuous activations," Neurocomputing, vol.123, pp.255–265, 2014.

[14] G.Zhang, J.Hu and Y.Shen, "New results on synchronization control of delayed memristive neural networks," Nonlinear Dynamics, vol.81, pp.1167-1178, 2015.

[15] X.Yang, J.D.Cao and W.Yu, "Exponential synchronization of memristive Cohen–Grossberg neural networks with mixed delays, "Cogn Neurodyn, vol.8, pp. 239–249, 2014.

[16] G.D.Zhang and Y. Shen, "Exponential synchronization of delayed memristor-based chaotic neural networks via periodically intermittent control," Neural Networks, vol.55, pp.1-10, 2014.

[17] A.Dai, W.Zhou and Y.Xu, "Adaptive exponential synchronization in mean square for Markovian jumping neutral-type coupled neural networks with time-varying delays by pinning control," Neurocomputing, vol.173, pp.809-818, 2016.

[18] Z.Cai, L.Huang L and L.Zhang, "New conditions on synchronization of memristor-based neural networks via differential inclusions, " Neurocomputing, vol.186, pp.235-250, 2016.

[19] J.N.Li, W.D.Bao and S.B.Li, "Exponential synchronization of discrete-time mixed delay neural networks with actuator constraints and stochastic missing data," Neurocomputing, vol.207, pp. 700-707, 2016.

[20] X.Han, H.Wu and B.Fang, "Adaptive Exponential Synchronization of Memristive Neural Networks with mixed time-varying delays," Neurocomputing, vol.201, pp.40-50, 2016.

[21] W.Zhang, C.Li and T.Huang, "Stability and synchronization of memristor-based coupling neural networks with time-varying delays via intermittent control," Neurocomputing, vol.173, pp. 1066-1072, 2016.

[22] A.Abdurahman, H.Jiang and Z.Teng, "Exponential lag synchronization for memristor-based neural networks with mixed time delays via hybrid switching control," Journal of the Franklin Institute, vol.353, pp.2859-2880, 2016.

[23] X.D.Li, R.Rakkiyappan and G.Velmurugan, "Passivity analysis of memristor-based complex-valued neural networks with time-varying delays," Neural Processing Letters, vol.42, pp.517-540, 2014.

[24] L.Duan and L.H.Huang, "Periodicity and dissipativity for memristor-based mixed time-varying delayed neural networks via differential inclusions," Neural Networks, vol.57, pp.12-22, 2014.

[25] Z.Y.Guo, J.Wang J and Z.Yan, "Global exponential dissipativity and stabilization of memristor-based recurrent neural networks with time-varying delays," Neural Networks, vol.48, pp.158-172, 2013.

[26] Z.Y.Guo, J.Wang and Z.Yan, "Attractivity analysis of memristor-based cellular neural networks with time-varying delays," IEEE Trans on Neural Networks and Learning System, vol.25, pp.704-717, 2014.

[27] M.Jiang, S.Wang and J.Mei, "Finite-time synchronization control of a class of memristor-based recurrent neural networks," Neural Networks, vol.63, pp.133-140, 2015.

[28] A.Chandrasekar, A.Rakkiyappan and J.D.Cao, "Synchronization of memristor-based recurrent neural networks with two delay components based on second-order reciprocally convex approach," Neural Networks, vol.57, pp.79-93, 2014.

[29] R.Rakkiyappan, G.Velmurugan and J.D.Cao, "Finite-time stability analysis of fractional-order complex-valued memristor-based neural networks with time delays,"Nonlinear Dynamics, vol.78, pp.2823-2836, 2014.

[30] N.Li and J.D.Cao, "New synchronization criteria for memristor-based networks: Adaptive control and feedback control schemes, " Neural Networks, vol.61, pp.1-9, 2015.

[31] M.D.Paola, F.P.Pinnola and M.Zingales, "Fractional differential equations and related exact mechanical models,"Computers and Mathematics with Applications, vol.66, pp. 608-620, 2013.

[32] V.Lakshmikantham and A.S.Vatsala, "Basic theory of fractional differential equations," Nonlinear Analysis Theory Methods and Applications, vol.69, pp.2677-2682, 2008.

[33] H.B.Bao and J.D.Cao, "Projective synchronization of fractional-order memristor-based neural networks," Neural Networks, vol.63, pp.1-9, 2015.

[34] G.Velmurugan, R.Rakkiyappan and J.D.Cao, "Finite-time synchronization of fractional-order memristor-based neural networks with time delays," Nonlinear Dynamics, vol.73, pp.36-46, 2016.

[35] G.Velmurugan and R.Rakkiyappan, "Hybrid projective synchronization of fractional-order memristor-based neural networks with time delays, " Nonlinear Dynamics, vol.11, pp.1-14, 2015.

[36] Bao H, Ju HP, Cao JD. Adaptive synchronization of fractional-order memristor-based neural networks with time delay [J]. Nonlinear Dynamics, 2015, 82:1-12.

[37] L.Chen, R.Wu and J.D.Cao, "Stability and synchronization of memristor-based fractional-order delayed neural networks," Neural Networks the Official Journal of the International Neural Network Society, vol.71, pp.37-44, 2015.

[38] Y.Gu, Y.Yu and H.Wang, "Synchronization for fractional-order time-delayed memristor-based neural networks with parameter uncertainty," Journal of the Franklin Institute, vol.353, pp.3657-3684, 2016.

[39] A.Abdurahman, H.Jiang and Z.Teng, "Finite-time synchronization for memristor-based neural networks with time-varying delays," Neural Networks, vol.69, pp.20-28, 2015.

[40] X.Yang, J.Cao and W.Yu, "Exponential synchronization of memristive Cohen-Grossberg neural networks with mixed delays," Cognitive Neurodynamics, vol.8, pp.239-249, 2014.

[41] J.Perez-Cruz, E.Portilla-Flores and P.Niño-Suarez, "Design of a nonlinear controller and its intelligent optimization for exponential synchronization of a new chaotic system," Optik -International Journal for Light and Electron Optics, vol.130, pp.201-212, 2016.

[42] Q.Gan, "Exponential synchronization of stochastic neural networks with leakage delay and reaction-diffusion terms via periodically intermittent control," Neural Processing Letters, vol.22, pp.393-410, 2013.

[43] S.Tyagi, S. Abbas and M.Kirane, "Global asymptotic and exponential synchronization of ring neural network with reaction-diffusion term and unbounded delay, " Neural Computing and Applications, pp.1-15, 2016.

[44] J.P.Aubin and A. Cellina, "Differential inclusions set-valued maps and viability theory," Acta Applicandae Mathematica, vol.6, pp.215-217, 1986.

Application of K-means Algorithm in Geological Disaster Monitoring System

Wang Jianguo
College of Computer Science and Engineering
Xi'an Technological University
No.2 Middle Xuefu Road, Weiyang District,
Xi'an, 710021, China
e-mail: wjg_xit@126.com,

Xue Linyao*
College of Computer Science and Engineering
Xi'an Technological University
No.2 Middle Xuefu Road, Weiyang District,
Xi'an, 710021, China
e-mail: 1525610807@qq.com

***Abstract*—The K-means algorithm is considered to be the most important unsupervised machine learning method in clustering, which can divide all the data into k subclasses that are very different from each other. As K-means algorithm is simple and efficient, it is applied to data mining, knowledge discovery and other fields. This paper proposes CMU-kmeans algorithm with improved UPGMA algorithm and Canopy algorithm. The experimental results is that the algorithm can not only get the number k of the initial clustering center adaptable, but also avoid the influence of the noise data and the edge data. Also, the improved algorithm can void the initial effect of the random selection on the clustering, which reflects the actual distribution in the dataset.**

Keywords-Clustering Analysis; CMU-kmeans Algorithm; Geological Disaster Monitoring Data

I. INTRODUCTION

The occurrence of geological disasters caused great casualties to humans, the main reasons include landslides and debris flow and rainfall and so on. And these geological disasters always cause many local public facilities to be damaged by large and small, and brought great damage to the people and their property. Also, there are still many such cases in China. Faced with such a severe threat of geological disasters, the state and the government on the prevention and control of geological disasters into a lot of human and material resources, and achieved remarkable results. With the progress of technology and high development of information technology, many new detection equipments have been put into the geological disaster real-time detection, such as GPS, secondary sound wave monitoring, radar and so on.

With the development of geological hazard detection technology, the amount of the monitoring data grew by leaps and bounds, data types are becoming more and more complex as well. K-means algorithm is a clustering algorithm based on the classification of the classic algorithm, the algorithm in the industrial and commercial applications more widely. As we all know, it both has many advantages and many disadvantages. In this paper, we mainly study the optimization of the initial clustering center and the avoidance of the blindness of the k-value selection, and propose the CMU-kmeans algorithm.

The data source of the study is the historical data detected by the geological disaster monitoring system, and 2000 records are randomly selected from the rainfall data of different areas in Shaanxi Province as the research object, which are served as a representative sample of the improved K-means clustering algorithm. The experimental results show that the improved algorithm not only eliminates the sensitivity to the initial input and improve the stability and effectiveness of the algorithm, but also can intelligently determine the initial clustering center number k, which improves the simplicity and operability of the algorithm.

A. Overview of K-means algorithm

The K-means algorithm is a classical unsupervised clustering algorithm. The purpose is to divide a given dataset containing N objects into K clusters so that the objects in the cluster are as similar as possible, and the objects between clusters are as similar as possible. Set the sample set X = {x1, x2, x3, ..., xn}, n is the number of samples. The idea of the K-means algorithm is that the k data objects are randomly selected from the sample set X as the initial clustering center, and then the data is allocated to the most similar cluster according to the similarity degree of each data object and k clustering centers; Recalculate the average of each new cluster and regard it as the next clustering center and repeat the process until the updated cluster center is consistent with the update, that is, the criterion function E converges. The goal is to make the object similarity in the cluster the largest, and the similarity between the objects is the smallest. The degree of similarity between the data can be determined by calculating the Euclidean distance between the data. For the n-dimensional real vector space, the Euclidean distance of two points is defined as form.1:

$$\delta(\xi, \psi)=\sqrt{(x_i - y_i)^2} \quad (1)$$

Here, x_i and y_i are the attribute values of x and y respectively, and the criterion function is defined as form.2:

$$E =\sum_{i=0}^{n} \sum_{x \in c_i} |x - \overline{x}_i|^2 \quad (2)$$

Here, k is the total number of clusters, and $\overline{x}_i$ is the center of cluster c. The flow of K-means algorithm is shown in Fig. 1.

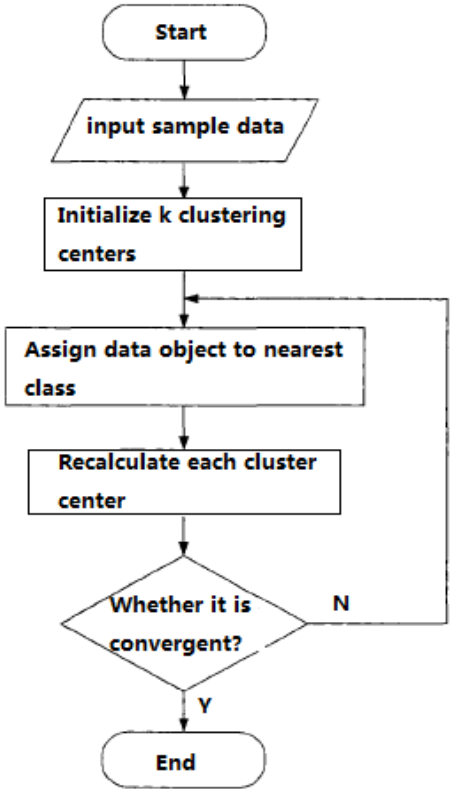


Figure 1. K-means clustering algorithm flow chart

B. Research status quo of K-means algorithm

For the advantages of K-means algorithm, it has been widely used in practice, but there are many shortcomings as well. In order to get better clustering effect, many researchers have explored the shortcomings of improving K-means. Aiming at the shortcomings of K-means algorithm in selecting the initial point, many scholars have proposed an improved method. Duan Guiqin [1] uses the method of product based on mean and maximum distance to optimize the initial clustering center. The algorithm first selects the set of data objects which are the farthest from the sample set to join the clustering center, and then the set of mean and current poly The largest data object of the class center is added to the clustering center set, which improves the accuracy. Yi Baolin [2] et al. proposed another improved K-means algorithm, which first calculates the density of the region to which the data object belongs, and then selects k points as the initial center in the high density region. The experimental results show that the algorithm reduces the initial center point Impact. Yiu-Ming Cheng[3] and others proposed a new clustering technique called K * -means algorithm. The algorithm consists of two separate steps. A center point is provided for each cluster in the first step; and then adjust the unit through adaptive learning rules in the second step. The algorithm overcomes the shortcomings of K-means algorithm initial center sensitivity and K value blindness, but the calculation is complicated. Xie and others [4] proposed a k-means algorithm to optimize the initial clustering center by using the minimum variance based on the sample space distribution compactness information. The algorithm chooses the samples with the smallest variance and a distance away from each other as the initial clustering center. Liu Jiaxing et al.[5] proposed a radius-based k-means + λ algorithm. When selecting the initial center point of the cluster, the distance ratio between points is calculated from the λ parameter and rounded at a specific distance. In the circle, an initialized center point is selected according to the distance ratio, and the algorithm has higher performance in error rate and operation time. Ren Jiangtao[6] proposed an improved K-means algorithm for text clustering, which is improved by using feature selection and dimension reduction, sparse vector selection, initial center point search based on density and spreading, Class accuracy, stability and other aspects have improved.

C. The performance analysis of K-means algorithm

K-means clustering algorithm uses the Euclidean distance to calculate the distance between each sample point. For the convex and spherical data distribution, the clustering effect is better and has been widely used in many fields. However, the Euclidean distance criterion adopted by the algorithm also has some limitations. For the more complicated or non-convex data, the clustering effect is often not very satisfactory. Clustering algorithm in the iterative process, if you do not meet the termination criteria will recalculate the average clustering center, this operation also improves the convergence rate of the clustering algorithm. In summary, K-means clustering algorithm has the following advantages and disadvantages of the following aspects.

1) The main advantages of K-means algorithm:

a) K-means clustering algorithm has high stability and scalability, clustering effect is very well.

b) The results of the treatment is intuitive and easy to understand. When dealing with the target data in numerical form, its geometric meaning is very clear. When clustering images and texts, the extracted eigenvalues can be regarded as clustering result values for the convenience of people's understanding.

c) K-means clustering algorithm When dealing with numerical data sets, the input data sequence will not affect the clustering result.

d) It can be a good judge of the data set shape is convex cluster.

2) The main shortcomings of K-means algorithm:

a) The K value in the K-means algorithm needs to be given in advance. According to the K value determined in advance, the clustering samples are classified into K class, so that the sum of squares of all the samples in the clustering domain to the clustering center is minimized.

b) Clustering results are highly dependent on the selection of initial clustering centers. The K-means algorithm uses the stochastic method to select the initial clustering center. If the initial clustering center is chosen improperly, it is difficult to obtain the ideal clustering effect. This dependence on the initial value may lead to the

instability of the clustering results, and it is easy to fall into the local optimal rather than the global optimal results.

c) Sensitive to noise and isolated points.

d) The time complexity of the algorithm is large.

II. IMPROVEMENT OF K-MEANS ALGORITHM AND ITS APPLICATION

Aiming at the shortcomings of traditional K-means algorithm, this paper mainly improves on the optimization of initial clustering center to enhance the clustering effect.

A. The selection of data object in Cluster analysis

The preliminary data are collected firstly when data selecting, then know about the characteristics of data to identify the quality of the data and to find a basic observation of the data or assume the implied information to monitor the subset of data of interest. The data object segmentation variable determines the formation of clustering, which in turn affects the correct interpretation of the clustering results, and ultimately affects the stability of the clustering clusters after the new data objects are added. Before the K-means clustering related data mining, the sample data set related to the data mining clustering analysis should be extracted from the original data object set, and it is not necessary to use all the historical data. In addition, we should pay attention to the quality of data, only high-quality data to the correct analysis of conclusions everywhere, to provide a scientific basis for clustering.

The source of this research object is the historical monitoring data of the geological disaster monitoring system. From the records of geological monitoring data from 2016 to 2017, a representative sample of K-means clustering algorithm for this improved algorithm is selected as the object of study in 2000, and the two samples of rainfall are randomly selected in different regions.

The sample data attributes show as table1:

TABLE I. THE SAMPLE DATA ATTRIBUTES

Field number	Field name	Field code	Type of data
1	Id	Xx	Number
2	Sno	Yy	Varchar
3	Type	type	Varchar
4	Gettime	time	Datatime
5	Alarm Level	alarm	Integer
6	Value	value	Double
7	Day Value	d_value	Double

For the cluster analysis, there are obviously redundant ones in the data attributes of the above geological hazard monitoring system, and it does not have the objectivity of the cluster analysis data. Therefore, the redundant ones should be eliminated. Finally, only four data object attributes reflecting the characteristics of rainfall data are selected as the research object. The optimized data attributes show as table2:

TABLE II. THE OPTIMIZED DATA ATTRIBUTES

Field number	Field name	Field code	Type of data
1	Id	xx	Number
2	Sno	yy	Varchar
3	Gettime	time	Datatime
4	Day Value	d_value	Double

B. Improvement of K-means algorithm

It is not difficult to see that, through the above study of the status quo, we can see that most of the above algorithm improvements are only a single defect in the traditional k-means algorithm is optimized. Although these improvements have optimized the k-means algorithm to some extent, there are still many shortcomings. For the above geological disaster monitoring system rainfall data characteristics, the K-means algorithm is very sensitive to the initialization center, and the initial clustering center is very easy to make the clustering result into the local optimum and the influence of the isolated point is large. In this paper, the simple random sampling technique is used to reduce the scale of the data set on the original dataset, and then the improved UPGMA algorithm and Canopy algorithm are combined to propose the CMU-kmeans algorithm. The improved algorithm can select the points with the furthest distance k in the high density region as the initial clustering center according to the regional density of each data, so that the improved k-means algorithm can produce high quality poly The results show that the sensitivity of the algorithm is not only eliminated, but also the stability and validity of the algorithm are improved.

1) Improved UPGMA algorithm

a) The basic idea of improved UPGMA algorithm

At the beginning of the UPGMA algorithm, each data object in the sample data set is considered as a separate class; and calculates the distance between each two data objects to obtains the distance matrix, then merges the two data objects that are closest to each other to obtain a new subclass, repeat the process .The UPGMA algorithm stops until no new class is generated or the stop condition is satisfied. It can be found that the first subclasses are usually located in the dense area of the data set, so the subclass center selected by this algorithm can be used as the initial clustering center candidate point for the next step. In this way, the selection of the initial clustering center is optimized and its accuracy is improved. The distance between two data objects is measured using the Euclidean distance formula, as form3:

$$\mathrm{d} = sqrt\left(\sum_{k=1}^{m} \left(X_{ik} - X_{jk} \right)^2 \right) \tag{3}$$

Here, Xi and Xj represent the data objects in the sample data set.

Xi={Xi1,Xi2,…,Xik,…,Xim}， k=1,2,…,m

Xj={Xi1,Xi2,…,Xik,…,Xjm}， k=1,2,…,m

The formula for calculating subclasses is as form4:

$$Z = \frac{1}{n}\sum_{j=1}^{n} X_j \tag{4}$$

Here, n refers to the number of data objects contained in a subclass, and Xj refers to a data object in the subclass.

b) The description of improved UPGMA algorithm

Input: All data in the sample data set, parameters m, p, Q;

Output: initial clustering center candidate point.

(1) set each data object as a separate class;

(2) Calculate the distance between two data objects, and then merge the nearest two classes into a new subclass to determine whether the subclass of the data object containing no less than m% of the total amount of data continues to produce , If not, then go to (4);

(3) For (i = 1 to maxcluster) {

{ for (j = i + 1 to maxcluster) {

If the number of data objects in subclasses i and j is less than or equal to m% of the total amount of data, calculate the distance between them to obtain the distance matrix.

}

}

Find the nearest two subclasses i and j and merge them into a new subclass, then add the new subclass to the end of the sequence Q to go to (2);

(4) Select the former p% subclasses in the sequence Q as the candidate subclasses and calculate the centers of all candidate subclasses as the initial clustering center candidate points.

Using the advantage of the improved UPGMA hierarchical clustering algorithm, we can find the dense region of the data set，which avoid the edge data and the noise data become the initial center candidate point. At the same time, considering the relative intensity of the region, we propose new clustering conditions and filter conditions to change the traditional UPGMA algorithm, so that the generation of subtrees can be stopped at different clustering levels to adapt to the actual density distribution data set. But the improved UPGMA algorithm also has some shortcomings. For example, if the m% and p% values are not set properly, the selection of the initial clustering center candidate points may be too dense and centralized. However, the Canopy algorithm, which introduces the idea of maximum and minimum distance, can select the data points that are far apart from each other. It is necessary to make up the deficiencies of the improved UPGMA algorithm. Therefore, it is necessary to introduce the Canopy algorithm to ensure that the distribution of the initial clustering center is decentralized, which can correctly reflect the data distribution of the original data set.

2) Improved Canopy algorithm

In order to avoid the clustering process is locally optimal, it is necessary to make Canopy get the center point spacing as large as possible. The maximum and minimum distance method [30] is a kind of test-based algorithm in the field of pattern recognition. Its basic idea is to take the object as far as possible as a cluster center, trying to get a better initial division. The algorithm not only intelligently determines the number k in the initial clustering, but also improves the efficiency of dividing the initial data set.

a) The description of improved Canopy algorithm

The Euclidean distance method is used to measure the degree of dissimilarity between data objects. Set the data set, S={X1,X2,…,Xn}, and the initial cluster center set is V = {v1, v2, ..., vn}. The improved Canopy algorithm is described as follows:

Input: Improve the initial clustering center candidate point of the UPGMA algorithm output, the parameter θ;

Output: Optimize the initial clustering center.

(1) Arbitrarily select a data object from the data set S as the first cluster center point v1 and put it into V;

(2) Calculate the distance between v1 and all the data objects remaining in the data set S, and put the farthest data object into V as the second cluster center v2;

(3) Calculate the distance Di between all the data objects Xi and all the data objects remaining in the data set S, select the smaller distance and denote Min (Di);

(4) Selects the maximum value in all the Min(Di) , marked as Max (Min (Di)), and regard the corresponding data Xi as the candidate cluster center, then judgment is made by the discriminant formula Max (Min (Di))> θ ‖ v1-v2 ‖. If the condition is satisfied, Xi is added to the initial clustering center set V, and if it is not satisfied,

(5) To (3);

(6) Output optimization of the initial clustering center.

The most critical step in the improved Canopy algorithm is the step (4), which takes the corresponding point of Max as the candidate of the new clustering center, thus avoiding the fact that the distance from an existing clustering center is closer to the other Clustering centers are far away as candidates for possible candidates. Therefore, the algorithm can be used to ensure that each new clustering center is far from the distance of the existing clustering center.

b) The analysis of advantages and disadvantages of improved algorithm of Canopy

The improved Canopy algorithm can use the k data objects farthest from each other in the data set as the initial clustering center, so as to avoid the situation that the initial clustering center distribution is too concentrated and intensive. But on the one hand, it is possible to select the noise data and the edge data, making the algorithm easy to fall into the local optimal solution, it is difficult to get the global optimal solution.

On the other hand, if the sample size of the whole data set is n, we need to scan the database first if we want to find a new cluster center each time; After finding the nearest distance from each object to the existing cluster center, we need scan the database to get the maximum-minimum distance. so we need a total of 2n distance calculation. The time complexity of the improved Canopy algorithm is: O (nk) .if the k clustering centers need to be found in the end of algorithm. Therefore, the computational complexity of the improved Canopy algorithm depends on the size of n, and there are thousands of objects in large databases usually, if we treat the original data set

with the improved Canopy algorithm directly, the implementation efficiency is low and the required storage space will be significantly increased.

3) MCU-kmeans algorithm

Generally, in order to ensure fully reflecting the distribution of data in the entire data set, every cluster center should be distributed in the high density area of the data set and dispersed as much as possible. Based on the above considerations, this paper proposes the MCU-kmeans algorithm, which combines the improved UPGMA algorithm and the improved Canopy algorithm to obtain the optimized initial clustering center, and then apply these optimized initial clustering centers to the k-means algorithm to enhance the clustering effect. Among them, the improved UPGMA algorithm is used to find the high density region, so that the selected initial clustering center candidate point away from the noise data and edge data; And the improved Canopy algorithm is used to avoid that the initial clustering center distribution is too concentrated and dense to ensure that the distances between the cluster center points are far away, which fully reflect the overall distribution of the data set. Therefore, the improved UPGMA algorithm and the improved Canopy algorithm complement each other so that the initial clustering centers selected by the algorithm are far apart from each other and all are located in the high density region of the data set. To sum up, the CMU-kmeans algorithm is as follows.

a) The initialization of the cluster center;

- Improved UPGMA algorithm: obtain the initial clustering center candidate point;
- Improved Canopy algorithm: obtain the appropriate initial clustering center;

b) K-means algorithm iteration;

c) The assessment of clustering results.

It can be seen that the framework of the CMU-kmeans algorithm is divided into three phases, as shown in Figure 2, the first stage of the algorithm is the initial optimization algorithm, which is the most important part of the improvement. The purpose is to intelligently capture the original The optimal initial clustering seed and the optimal initial clustering number of the data set distribution. The second stage is the main body of the algorithm, and the K-means algorithm is used to cluster on the whole data set and get the clustering result. The third stage is experiment and evaluated to verify the validity of the proposed CMU-kmeans algorithm.

It can be seen that the framework of the CMU-kmeans algorithm is divided into three phases, as shown in Fig.2 , the first stage of the algorithm is the initial optimization algorithm, which is the most important part of the improvement. The purpose is to intelligently capture the optimal initial clustering seed and the number of the data set distribution. The second stage is the main body of the algorithm, and the K-means algorithm is used to cluster on the whole data set and get the clustering result. The third stage is experiment and evaluated to verify the validity of the proposed CMU-kmeans algorithm.

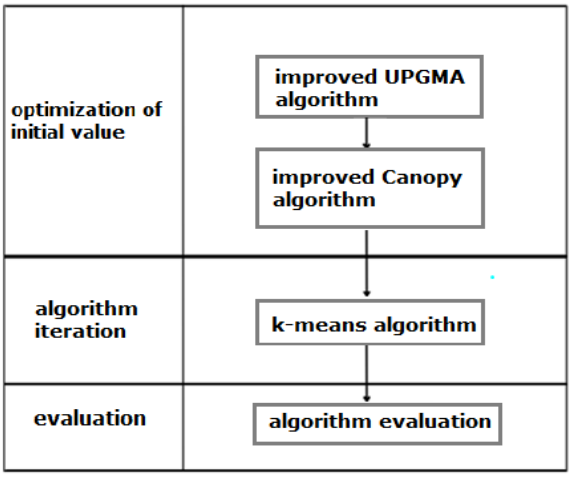


Figure 2. CMU-kmeans algorithm framework

The CMU-kmeans algorithm proposed in this paper can effectively reduce the dependency of the k-means algorithm on the initial clustering center selection. For the data set with uneven data distribution, on the one hand, it avoids the idea that the initial clustering center is too dense; On the other hand, it avoids the fact that the selected initial clustering centers are too scattered and even select noise data and edge data is happening, which can improve the stability and validity of the algorithm. At the same time, the number k of the initial clustering center can be automatically determined without the pre-set and the simplicity and maneuverability of the algorithm can be improved.

III. EXPERIMENT ANALYSIS

A. Experimental description

The data set selected from the experiment comes from the rainfall data collected in the geological hazard detection system and the rainfall data set after the artificial noise is added. The experimental environment is: Inter(R)Core(TM)i3-2330M,4G RAM，250G hard disk，Win7 operating system.

In order to verify the validity and stability of the algorithm, the traditional K-means clustering algorithm, the improved Canopy algorithm and the CMU-kmeans algorithm are compared under the rainfall data set. The clustering result of the traditional k-means algorithm is an average of 10 executions. Evaluate the performance of the algorithm according to the accuracy of the clustering results and the recall rate.

B. Performance evaluation criteria

The traditional k-means algorithm, the improved Canopy algorithm and the clustering effect of CMU-kmeans algorithm proposed in this paper are evaluated by the commonly used evaluation method to evaluate the quality of clustering effect, namely, precision and recall. The accuracy and recall rate are defined as follows:

$$P（i, j） = \text{precision}(i, j) = Ni,j / Ni \quad (5)$$

$$R（i, j）= recall(i, j) = Ni,j / Nj \quad (6)$$

Here, Ni, j represents the number of classes i in cluster j; Ni is the number of all objects in class i; Nj is the number of all objects in cluster j.

C. *Experimental content and structure analysis*

Table3 below shows the detailed experimental results of the three algorithms on the geo-disaster monitoring system rainfall data set.

TABLE III. DETAILED EXPERIMENTAL RESULTS ON THE RAINFALL DATASET

Rainfall set	k-means algorithm		Improved Canopy algorithm		CMU-kmeans algorithm	
	precision	recall	Precision	recall	precision	recall
1st	25.423	26.125	50.799	61.078	56.939	65.783
2nd	24.287	25.365	52.975	63.288	56.423	64.921
3rd	25.61	18.864	48.895	58.887	57.413	66.174
4th	27.143	26.143	53.425	63.683	57.682	68.108
5th	22.365	25.31	50.073	58.404	56.163	65.063
6th	18.102	26.421	50.444	64.65	56.468	65.224
7th	25.326	24.623	49.362	57.338	58.921	67.638
8th	28.325	26.852	49.975	60.075	56.239	66.405
9th	26.562	28.154	54.267	62.392	58.341	66.423
10th	23.985	26.523	51.445	60.651	57.267	65.392
average	25.013	25.938	51.666	61.045	57.186	66.113

As can be seen from the above table, in ten experiments, values of the two performance evaluation criteria (precision and recall)vary greatly based on the traditional k-means algorithm ,showing a very unstable state. To precision as an example, the minimum value of the ten experimental results is 18.102, and the maximum is 28.325, the difference is 10.323, and the recall is different from 9.290. The result of the improved Canopy algorithm has improved, the precision is 6.549, and the difference is 6.345.

In the CMU-kmeans algorithm, the values of the two performance evaluation criteria are obviously improved and are still stable. The precision of the ten results is 56.163, the maximum is 58.921, the difference is 2.758, and the recovery value is 3.187.

In order to make the experimental results more straightforward, the above 10 experimental results with the wave diagram shown in order to compare the stability of the two algorithms and accuracy.

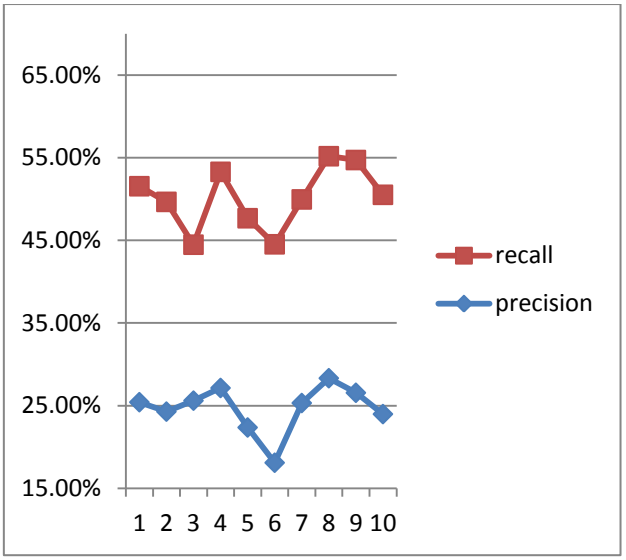


Figure 3. The precision and recall values of the traditional k-means algorithm

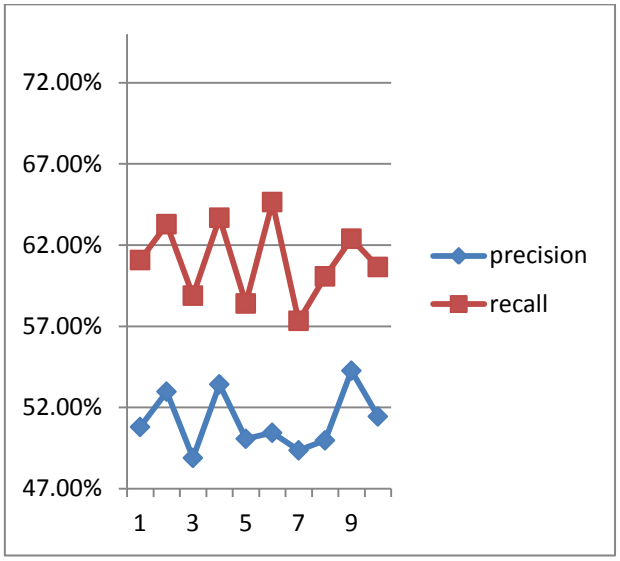


Figure 4. The precision and recall values of the improved Cannopy algorithm

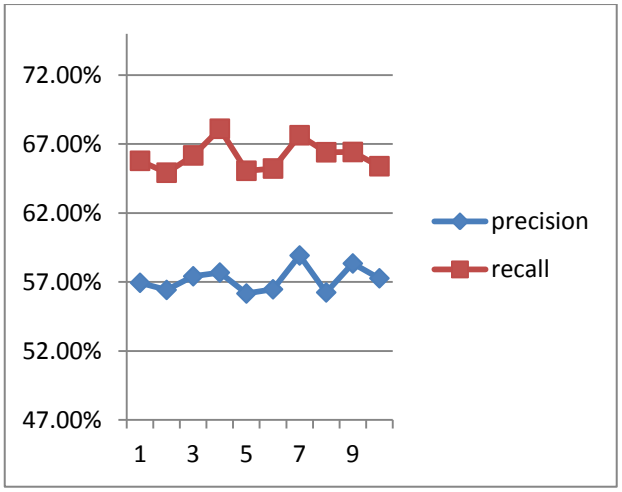


Figure 5. The precision and recall values of the CMU-kmeans algorithm

It can be seen from Fig.3 , Fig.4 and Fig.5 that the improved Canopy algorithm clustering effect is improved on the basis of the traditional k-means algorithm. The improved CMU-kmeans algorithm can improve the state of the improved Canopy algorithm. The precision and recall value of the CMU-kmeans algorithm are small and obviously improved, and the lifting effect is significant.

IV. CONCLUSION

The CMU-kmeans algorithm improves the clustering effect, make the performance tend to be stable, and the computational complexity of the calculation is obviously reduced compared with the traditional k-means algorithm and the improved Canopy algorithm. Also, the algorithm can adaptively determine the number k of the initial clustering center, avoid the influence of the noise data and the edge data and random selection of initial clustering center, and also well reflect the actual distribution of clustering center in the dataset.

REFERENCES

[1] Zhai D H, Yu J, Gao F, et al. k-means text clustering algorithm based oninitial cluster centers selection according to maximum distance [J]. Application Research of Computers, 2014, 31(3):379 – 382.

[2] Baolin Yi, Haiquan Qiao, Fan Yang, Chenwei Xu. An Improved Initialization Center Algorithm for K-Means Clustering[C]. Computational Intelligence and Software Engineering, 2010, pp:1-4.

[3] Redmond S J, Heneghan C.A method for initializing the K-means clustering algorithm using kd-trees[J]. Pattern recognition letters, 2007, 28(8):965-973.

[4] Liu J X, Zhu G H, Xi M. A k-means Algorithm based on the radius [J]. Journal of Guilin University of Electronic Technology,2013,33(2):134-138.

[5] Habibpour R, Khalipour K. A new k-means and K-nearest-neighbor algorithms for text document clustering [J]. International Journal of Academic Research Part A, 2014,6(3) : 79 — 84.

[6] Data mining techniques and applications - A decade review from 2000 to 2011[J]. Shu-Hsien Liao, Pei-Hui Chu, Pei-Yuan Hsiao. Expert Systems With Applicati--ons . 2012 (12).

[7] Application of Improved K-Means Clustering Algorithm in Transit Data Collection. Ying Wu, Chun long Yao. 20103rd International Conference on Biomedical Engineering and Informatics (BMET) . 2010.

[8] Zhou A W, Yu Y F. The research about clustering algorithm of K-means [J]. Computer Technology and Development, 2011,21(2):62-65.

[9] Duan G Q. Auto generation cloud optimization based on genetic algorithm [J]. Computer and Digital Engineering, 2015,43(3):379-382.

[10] Wang C L, Zhang J X. Improved k-means algorithm based on latent Dirichlet allocation for text clustering [J]. Journal of Computer Applications,2014,34(1):249-254.

[11] Deepa V K,Geetha J R R. Rapid development of applications in data mining[C]. Green High Performance Computing (ICGHPC),2013,pp:1-4.

[12] Sharma S, Agrawal J, Agarwal S, et al. Machinelearn-ing techniques for data mining: A survey[C]. Com-putational Intelligence and Computing Research (ICCIC),2013,pp:1-6.

[13] Efficient Data Clustering Algorithms: Improvements over Kmeans[J]. Mohamed Abubaker, Wesam Ashour. International Journal of Intelligent Systems and Applications(IJISA). 2013 (3).

[14] Fahad A, Alshatri N, Tari Z, Alamri A. A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis[C]. Emerging Topics in Computing.2014:267-279.

[15] Abubaker M, Ashour Wesam. Efficient data clustering algorithm algorithms:improvemwnts over k-means[J]. I nternational Journal of Intelligent Systems and Applications.2013(3):37-49.

[16] Tang Zhaoxia, Zhang Hui. Improved K-means Clustering Algorithm Based on Genetic Algorithm[C], Telkomnika Indonesian Journal of Electrical Engineering.2014, pp:1917-1923.

Research on Multi - Resonant LCL Harmonic Suppression Strategy

Jingwen Chen, Xin Zhou and Hongshe Dang*
School of electrical and information engineering
Shaanxi University of Science and Technology
Shaanxi 710021, China
*Address correspondence to this author at xuefu Road,
Xi'an, China, 710021;
e-mail: 2452744621@qq.com

***Abstract*—Aiming at the resonance problem in the process of grid connection of LCL filter microgrid inverter, a multi-resonance LCL harmonic suppression strategy is proposed. On the basis of analyzing the principle and establishing the mathematical model in detail, the realization process of the multi-resonance constant power compound control strategy is studied emphatically. Through the simulation, the validity of the control strategy is verified, The results show that the scheme stabilizes the output power and reduces the total harmonic distortion of the grid-connected inverter to 0.12%, and the corresponding phase current distortion rate drops to 0.02%.The suppression effect is obvious, it is an effective harmonic suppression method.**

Keywords-Microgrid Inverter;Harmonic;Multi Resonance Control;Constant Power Control

I. INTRODUCTION

With the depletion of traditional energy sources, the new energy power generation system with microgrid as the carrier has been developed rapidly because of its flexible, decentralized, small, close to users and the use of clean energy. Due to the energy structure of the micro grid mainly clean energy such as wind power, photovoltaic power generation, the distributed energy will generally need electricity to the grid through power electronic converter device to realize grid connected. Therefore, a lot of power electronic devices access to power grid harmonics, caused the converter power factor lower and parallel resonant circuit or series resonance, decrease active reactive power measurement accuracy, reduce the quality of power supply a series of problems, give the user the safe and security, economic operation of power system brings great harm. so harmonic suppression is very important.

Current research of harmonic suppression methods, mainly has: the harmonic suppression method based on active filter [1] [2] [3]; the micro-grid harmonic suppression based on virtual impedance [4]; LCL type Grid-connected inverter harmonic suppression [5] [6] [7] [8] and so on. Compared to L-filter, LCL-type filter has a third-order low-pass filter characteristics,(LCL filter with third order low pass filter properties), so for the same harmonic standard and lower switching frequency, we can use a relatively small filter inductor design, effectively reduce the system size(volume) and reduce losses, but the same will bring resonance problems. In this paper, a harmonic control strategy of micro-grid inverter based on PI control, multi-resonance control and LCL constant power control is proposed, which is used in the process of grid-connected control of micro-grid inverter to further reduce and net voltage of the total harmonic distortion rate, get better power of the grid.

II. MATERIAL AND METHODS

A. *Principle block diagram of multi resonance LCL harmonic suppression strategy*

Figure 1 is LCL multi-resonant constant power grid control system block diagram. Where P_{ref} Q_{ref} are the

actual active and reactive power reference values, v_{abc}、i_{abc} are the actual values of the grid voltage and current, v_d、v_q and i_d、i_q are the voltage components of the dq axis, i_{dref} and i_{qref} are the capacitor current reference Value, i_{cd}、i_{cq}、i_{cdref}、i_{cqref} are the capacitance current detection value on the dq axis components and capacitance current reference value.

The figure includes PQ control module, current loop multi-resonance control module, and PWM modulation module etc. Micro-network inverter grid output voltage v_{abc} and current detection values i_{oabc} ,after αβ coordinate transformation to get v_α、$v_\beta$and i_α、i_β,will be sent to the PQ controller; PQ controller according the active and reactive setpoint P_{ref} and Q_{ref} to calculate the current reference values$i_{\alpha ref}$ and$i_{\beta ref}$ and then compare with the current detection value of the αβ components $i_\alpha$and i_β. And after the ratio multi-resonant regulator $G_2(S)$ in the current loop control module, the reference values $i_{c\alpha ref}$, $i_{c\beta ref}$ of the capacitive current are obtained, then compared with the capacitance current detection value αβ component $i_{c\alpha}$and $i_{c\beta}$. And then adjusted by the proportional regulator G_1 (S), the control PWM circuit drives the inverter, so that the inverter output active power and reactive power constant. In order to meet the requirements of system stability, the current loop control module in Figure 1 solves the resonance problem caused by LCL, and achieves the purpose of suppressing low frequency harmonics and improves the system accuracy.

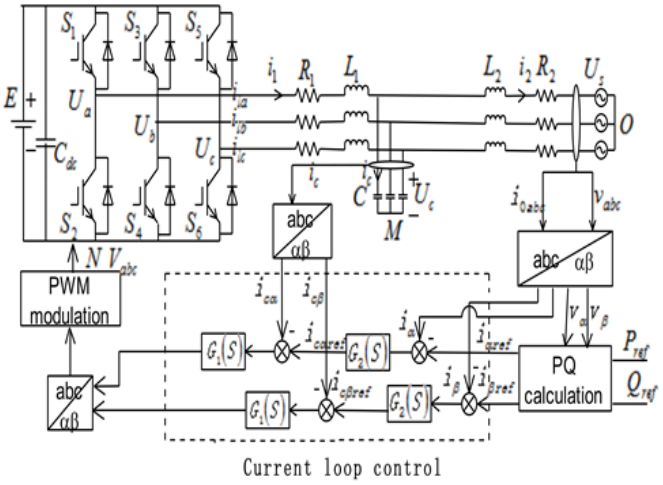


Figure 1. Principle block diagram of multi resonance LCL harmonic suppression strategy

B. Multi - resonant LCL Grid - connected Control Mathematical Model

Figure 1 constitutes a third-order LCL-type filter, where L_1 is the inductance, R_1 is its internal resistance and the equivalent resistance between the upper and lower legs of each phase, R_2 is the internal resistance of L_2. In the case of three-phase grid voltage symmetry, the mathematical model is as follows:

$$L_1\frac{di_1(t)}{dt}+R_1i_1(t)+L_2\frac{di_2(t)}{dt}+R_2i_2(t) = U_{dc}(t)S_k(t)+U_{NO}-U_s(t) \tag{1}$$

$$C\frac{du_c(t)}{dt}=i_1(t)-i_2(t) \tag{2}$$

$$i_{dc}(t)=C_{dc}\frac{du_{dc}(t)}{dt}+\sum_{k=a,b,c}i_k(t)S_k(t) \tag{3}$$

$$U_{NO}(t)=-\frac{u_{dc}(t)}{3}\sum_{k=a,b,c}S_k(t) \tag{4}$$

Where S_k is the switching function of the power switching device, when $S_k = 0$, the upper arm is turned on and the lower arm is turned off; When $S_k = 1$, the upper arm off, the lower arm conduction. Corresponding to the relationship between α and β stationary coordinate system is as follows:

$$\begin{aligned}
L_1 \frac{di_{1\alpha}}{dt} &= u_\alpha - u_{ca} - R_1 i_{1\alpha} \\
L_1 \frac{di_{1\beta}}{dt} &= u_\beta - u_{c\beta} - R_1 i_{1\beta} \\
L_2 \frac{di_{2\alpha}}{dt} &= u_{c\alpha} - e_\alpha - R_2 i_{2\alpha} \\
L_2 \frac{di_{2\beta}}{dt} &= u_{c\beta} - e_\beta - R_2 i_{2\beta} \\
i_{c\alpha} &= i_{1\alpha} - i_{2\alpha} \\
i_{c\beta} &= i_{1\beta} - i_{2\beta}
\end{aligned} \tag{5}$$

$i_{1\alpha}$、$i_{1\beta}$、$i_{2\alpha}$、$i_{2\beta}$ are the α and β components of the input and output currents in the αβ coordinate system; $i_{c\alpha}$、$i_{c\beta}$、$u_{c\alpha}$、$u_{c\beta}$ are the α and β components of the capacitive current and voltage in the αβ coordinate system; e_α and e_β are the α and β components of the grid voltage in the αβ coordinate system.

C. A block diagram of current loop control

LCL-type filter in the better suppression of high-frequency harmonics at the same time, because of its own structure for a third-order system, easy to produce resonance, the frequency near the narrow band and too high gain, will lead to the system and the load Parameter changes are very sensitive, affecting the stability of the system to bring a series of impact and harm to the grid. In order to reduce its sensitivity and high gain characteristics, to achieve the AC signal without static tracking, this paper on the basis of the use of active damping introduced into the capacitor current loop regulation to suppress high frequency interference, and the external loop current using proportional resonance control , Constructs a transfer function that performs AC compensation on the reference input signal. So that in a specific bandwidth in the same frequency response characteristics, to meet the system stability requirements, so that the output at the resonant frequency at high gain, the other frequency segment attenuation. Thus reducing the resonance, improve the stability of the system and control accuracy.

The control block diagram is showas follows:

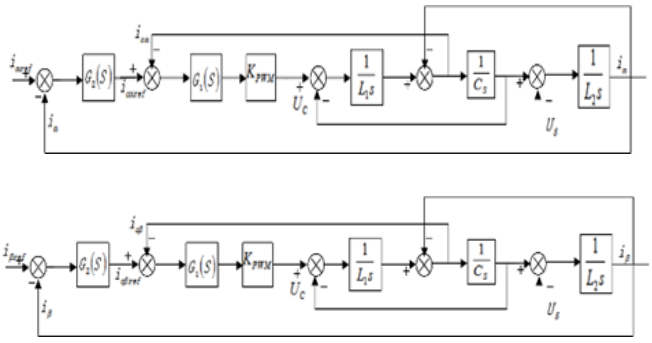


Figure 2. Block diagram of current loop control

As shown in Figure 2 where $G_2(S)$selected proportional resonance regulator, $G_1(S)$selected proportional regulator. After the current is transformed by the coordinate, the voltage and current v_α、v_β and i_α、i_β in the two stationary coordinates are obtained and sent to the PQ calculation module to obtain the reference current $i_{\alpha ref}$、$i_{\beta ref}$,and then compared with $i_\alpha$andi_β obtained the deviation by the proportionmulti-resonant regulator $G_1(S)$ in the current loop control module , get the capacitor current reference value $i_{c\alpha ref}$and $i_{c\beta ref}$, And it is compared with the capacitance current detection value αβ component $i_{c\alpha}$and $i_{c\beta}$, after adjusting the proportional regulator $G_2(S)$, then control PWM circuit drives the inverter, so that the inverter output active power and reactive power constant.

D. PR control

Since the PR regulator is equivalent to the PI modulator in the stationary coordinate system under the αβ coordinate,

the PR regulator can also be used to design the PI regulator parameter. Figure 2 in the parallel current and reference current deviation, through the multi-resonant control get the capacitor current reference value i_{cref} , i_{cref} and the actual capacitance of the current deviation, and then through the proportional control, the resulting signal through the PWM modulation to achieve active damping control. The use of a proportional feedback control of the capacitor current, stabilize the capacitor voltage, and enhance the stability of the system. The parallel current and capacitive current transfer functions are as follows:

$$G(S)=\frac{I_2(S)}{I_{ref}(s)} = \frac{KK_{PWM}K_\rho s+KK_{PWM}K_i}{L_1L_2Cs^4+KK_{PWM}L_2Cs^3+Ls^2+KK_{PWM}s+KK_{PWM}K_i} \quad (6)$$

The corresponding characteristic equation is:

$$D(S)=L_1L_2Cs^4+KK_{PWM}L_2Cs^3+Ls^2 +KK_{PWM}s+KK_{PWM}K_i \quad (7)$$

According to the Rouse stability criterion, the system stability condition is calculated as:

$$K_\rho<1+L_2/L_1$$
$$K_\rho(L-K\rho L_1)-KK_{PWM}K_iL_2C>0 \quad (8)$$

According to the stability criterion (formula 9) to set the parameters, so that when the grid to reach a stable state.

III. MULTIPLE PR CONTROL

A single PR regulator generates an infinite gain at a specific frequency. In order to ensure its stability and easy to achieve, using the approximate structure; its transfer function is as follows:The transfer function is as follows:

$$G_{PR}=K_P+\frac{2K_hw_cs}{s^2+2w_cs+w_h} \quad (9)$$

Where K_P is the scale factor, w_c is the frequency adjustment coefficient, $K_h$is the resonance coefficient, and w_h is the resonant frequency.

In order to achieve the 5,7,11 harmonic current compensation need to re-connect three resonant controller.

The transfer function of the current inner loop multi-resonance controller is:

$$G_{PR}=K_P+\sum_{i=5,7,11}\frac{2K_hw_cs}{s^2+2w_cs+w_h} \quad (10)$$

The minimum value of the resonant frequency in the LCL grid-connected inverter is:

$$f_r=\frac{1}{2\pi}\sqrt{\frac{L_1+L_2}{L_1L_2C}} \quad (11)$$

Then the minimum value of K is:

$$K_{min}=\frac{2L_1}{(L_1+L_2)E} \quad (12)$$

In order to ensure the stability of the control system, the cutoff frequency should be chosen to be less thanf_r so the K_P in the multi-resonance PR control can be approximated by the cutoff frequency f_c:

$$K_P=\frac{4\pi f_c(L_1+L_2)}{E} \quad (13)$$

According to the scope of K and the specific control requirements, through the control system open-loop baud diagram for parameter adjustment.

IV. DISCUSSION

Based on the detailed analysis of the principle of multi-resonance constant power control, the simulation results of the mathematical model are as follows

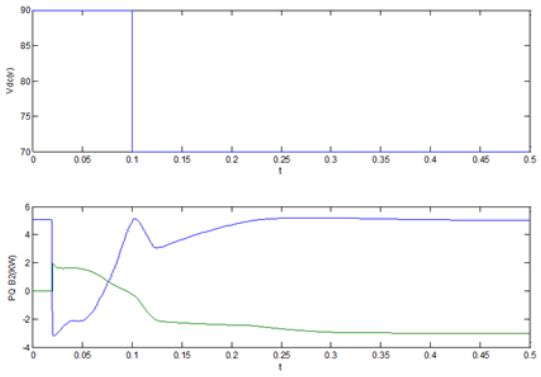


Figure 3. DC side voltage control after active, no power waveform

As shown above, the active and reactive power of the output after LCL multi-resonant constant power control is constant, which ensures the stable operation of the system.

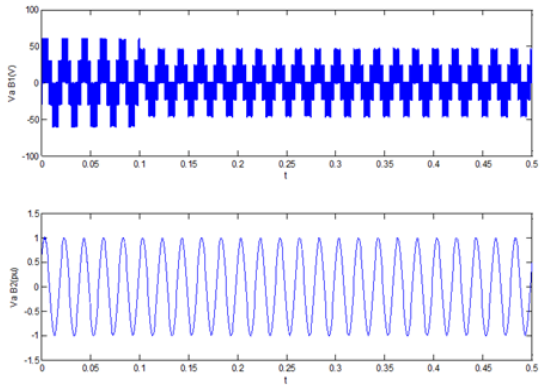


Figure 4. By LCL multi-resonant constant power control before and after the voltage waveform

From Fig. 4, we can find that the voltage and current waveforms before the LCL multi-resonant constant power control are unstable and distorted. After the control of the voltage and current waveform is improved, harmonic suppression effect is obvious.

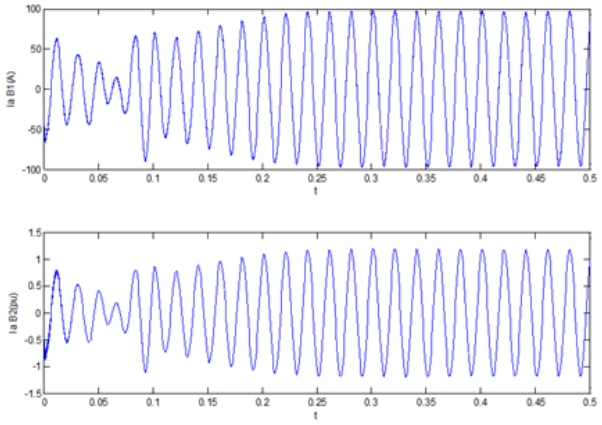


Figure 5. By LCL multi-resonant constant power control before and after the current waveform

From the above figure can be found by the LCL multi-resonant constant power control after the current waveform has improved.

In order to analyze the filtering effect by LCL multiresolution constant power control, the voltage distortion rate, the total voltage distortion rate and the current distortion rate of the respective voltage waveforms before and after the control are summarized as shown in Table 1 and 2 respectively.

TABLE I. LCL MULTI-RESONANT CONSTANT POWER GRID-CONNECTED CONTROL BEFORE AND AFTER THE VOLTAGE HARMONIC CONTENT

	Voltage distortion rate	**current distortion rate**
Before filtering	76.03%	15.23%
After filtering	0.12%	0.02%

TABLE II. LCL MULTI-RESONANT CONSTANT POWER CONTROL BEFORE AND AFTER THE VOLTAGE AND CURRENT DISTORTION

Number of harmonics	5	7	11
Before filtering	2.86%	0.93%	0.98%
After filtering	0.08%	0.04%	0.01%

From Table 1, it can be found that the contents of the harmonics before the filtering are reduced and have an inhibitory effect.

Table 2 can be seen by LCL multi-resonant constant power control filter before the voltage distortion rate of 76.03%, a phase current distortion rate of 15.23%, filtered voltage distortion rate reduced to 0.12%, the corresponding phase current distortion Small to 0.02 inhibitory effect is obvious.

V. CONCLUSION

In this paper, a harmonic suppression strategy for micro-grid inverter combined with LCL constant power control and multi-resonant PI control is proposed for the resonant problem of LCL filter microgrid inverters. It is found that the scheme stabilizes the output power LCL and reduces the total harmonic distortion rate of the grid inverter to 0.12% and the corresponding phase current distortion rate is reduced to 0.02%. The harmonic suppression effect is obvious.

REFERENCES

[1] Shang taohong. Study and parameter design of shunt hybrid active power filter [J] proceedings of the computer science, 2012, 29 (11): 327-330.

[2] Li Yan, Luo an, Fang Lu, Wang Wen. High voltage type hybrid active power filter [J]. Journal of electric technology, 2013, 28 (6): 147-157

[3] Wang Jidong, Qin Meicui. Micro grid harmonic suppression method [J]. Journal of Tianjin University APF based on the (Science and Technology Edition), 2015, 48 (7): 637-642

[4] Li Li. Study on hierarchical control and power quality improvement of microgrid [D]. Beijing: North China Electric Power University, 2015

[5] Han Yongru, Xue Shilong, and Study on the control strategy of grid connected inverter based on LCL filter [J]. Shanghai: Maritime University, 2015, 1 (103-109): (in Chinese)

[6] Huang Yafeng, Li Long, Yan Gangui large capacity PV inverter LCL filter parameter optimization design [J]. Beijing: North China Electric Power University, 2013, 41 (21): 104-110

[7] Zhang Xing, Li Fei, and in the grid connected inverter LCL filter improved topology [J]. Hefei: HeFei University of Technology, 2014,6:10-17

[8] Xu Jinming, Xie Shaojun, L L filter grid connected inverter robust current control [J]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012,36 (19): 99-104

[9] Yang Kun, Xie Chuan, Chen Guozhu. Current control of static reactive power generator based on frequency adaptive resonance controller [J]. Journal of electrical engineering, 2014, 29 (8): 249-254.

[10] Zhang Zhicheng, Liu Zhenlai, Guan Huchang. Island micro grid harmonic control method of [J]. power electronic technology, 2015, 49 (12): 135-13

Chen Jingwen (1978-), male (Han nationality), Inner Mongolia Chifeng people, associate professor, graduate tutor, research direction for micro-grid technology and application, E-mail: chenjw@sust.edu.cn

Zhou Xin (1993-), female (Han), Xianyang, Shaanxi, Master, research direction for micro-grid harmonics, inter-harmonics of the study

Dang hongshe(1962-), male (Han nationality), ShaanXiXianyang people, professor, research direction for micro-gridtechnology and application, (Corresponding author: E-mail: 2452744621@qq.com)

Searchable Re-encryption Cloud Storage Method Based on Markov Chain

Wang Hui[a], Wang ZhongSheng[b]
School of Computer Science and Engineering,
Xi'an Technological University,
Xi'an, 710021, China
e-mail: [a]277019826@qq.com;
[b]59483672@qq.com

Li Jinguang
Department of Information Technology
Shaanxi Heavy Duty Automobile CO.LTD
Xi 'an, 710200, Shaanxi, China

***Abstract*—Cloud storage is an emerging paradigm that offers on-demand, flexible, and elastic computational and storage services for the terminal users. When the large amount of data increases dramatically, the storage efficiency of the system would be decreased seriously. In this paper a new method of SReCSM(Searchable Re-encryption Cloud Storage Method) based on Markov chain is proposed. It predicts periodically by using the steady Markov strategy in stages and easy to select the optimal storage node. The data is scheduled to store in the node with the lowest cost in real time, and the node is selected to implement cloud storage access on mobile terminal. By using searchable re-encryption method, SReCSM has increased the storage requirement flexibility and minimize cost and searching time. And then the reliability model of SReCSM is established. Simulation results show that SReCSM introduced in this paper has the ability to predict accurately when the size of the data is different. Moreover, the influence of storage efficiency is reduced effectively through SReCSM when different size of the data is stored in storage nodes regardless of the storage cost. It is verified that the SReCSM based on Markov chain has higher reliability.**

Keywords-Cloud Storage; Markov Chain; Re-encryption; Reliability Model

I. INTRODUCTION

In order to meet the various storage demands, cloud storage is designed to store data in cloud and is widely used in the Internet. Compared with traditional data storage, it greatly improves the efficiency of the mass data storage and utilization of network resource. However, access from a mobile device to data, stored in a cloud, leads to poor client quality experience [1-3]. It is essential to reduce user download wait time for a requested file from a network to enhance client quality experience. As a result, cloud storage techniques are quite challenging. On one hand, storage density of the cloud storage is not big and the comprehensive storage efficiency is low. On the other hand, the high latency limited the use of mobile cloud storage, especially for the applications with frequent random accesses to a large set of small files. Therefore, cloud storage faces serious security and efficient problems [4-6].

Traditional cloud storage systems do not adapt well to different application environment and does not guarantee the integrity and confidentiality of cloud data. In other words, the cloud storage service does not guarantee that the data and operation of mobile users will not be lost, damaged, leaked, or illegally exploited by malicious or nonmalicious. Therefore, it's very dangerous for sensitive data to be stored directly in the cloud. The reliability of the mobile cloud storage depends on the extent of the impact on system storage efficiency while the storage solution fails [7]. Therefore, storing sensitive data on untrusted server is a challenging issue [8]. Simple encryption techniques have key management issues and which can't support complex requirements such as query, parallel modification, and fine-grained authorization. To guarantee confidentiality and proper access control of sensitive data, classical encryption are used [9-10].

To solve the problems brought by the hysteretic and density of tranditional storage methods in the cloud storage system, in this paper, SReCSM, Searchable Re-encryption

Cloud Storage Method is proposed using the idea of Markov chain. In this method, instead of using traditional storage, a proactive storage approach is adopted to avoid delayed storage. In addition, the SReCSM predicts periodically to increase the effiency of cloud storage. The major contribution of our work includes:

- Searchable Re-encryption Cloud Storage Method based on Markov chain is proposed and built as a SReCSM model. In order to find out the optimal storage policy in the cloud, an optimization problem is formulated as a Markov chain. The SReCSM model is designed to predict periodically by using the steady Markov strategy in stages and determine the lowest storage cost by which the storage benefit can be maximized.
- A value iteration algorithm of SReCSM is presented for computing the stationary deterministic policy. The data is scheduled to be stored in the node with the lowest storage cost in real time. Therefore, the optimal storage node must be selected to implement cloud storage access on mobile terminal.
- The central idea of the searchable re-encryption is proposed, which is to generate a re-encryption key and decrypt while the keyword matched. The objective of the proposed searchable re-encryption method is to increase the storage requirement, flexibility and reduce the security issues, overhead ratio and minimize the cost and searching time.
- The reliability of SReCSM model is proposed and analyzed. Because of the large scale of the cloud storage system, method of Monte Carlo simulation is adopted in the reliability evaluation of the cloud storage system.
- We conduct a series of experiment to validate the effectiveness, performance and robustness of SReCSM. Results show that the present approach can store mass data in cloud system effectively and securitily.

The rest of the paper is organized as follows. In section 2, related work is discussed. Section 3, Searchable Re-encrypption Cloud Storage Method based on Markov Chain is proposed. Section 4 describes the searchable re-encryption method in SReCSM. Section 5 presents the basic prototype system of SReCSM and simulation experiments. In addition, the security, performance and robustness of SReCSM are analyzed. Section 6 concludes the paper.

II. RELATED WORK

Efforts have been taken by researchers, developers, practitioners, and educators to identify and discuss the technical challenges and recent advances related to cloud storage. Han et al. proposed [11] multi-path data prefetching in mobile cloud storage. Multi-Path prefetching tend to prefetch more successors to ensure high prefetching hit rate and efficiency, and achieve a higher overall throughput performance of accessing cloud storage services via mobile network. Based on cloud storage, Lee et al. [12] designed an efficient delta synchronization (EDS) algorithm for mobile cloud storage applications, which can aggregates the updated data to reduce the synchronization traffic and synchronizes the aggregate done periodically to satisfy the consistency. Wang et al. [13] presented an Optimized Replica Distribution Method (ORDM) in cloud storage system. Zhang et al. [14] conducts modeling analysis of a cloud storage system, and proposes a Markov decision process modeling framework to analyze the reliability of the storage system. Chen et al. [15] proposed a new metric called joint response time, which not only considers the waiting time when the requested data are unavailable but also the queuing delay and service time when data become available. These methods mentioned above achieve cloud storage, but the cloud storage according to data size based on Markov is not considered. Aiming at the problem that the cloud storage according to data size based on Markov, SReCSM is proposed in this paper to satisfy the cloud storage periodically.

Most of the existing security schemes that are designed for mobile cloud storage environment are based on traditional cryptographic methods or pairing-based cryptography. The further deployment of cloud storage is limited by its security risks. The schemes presented in

[17-19], and [20, 25] provided the security features for mobile user in the cloud environment using the traditional cryptography methods. The rest of the security schemes discussed in [21-23], and [24] are based on pairing based cryptography for secure offloading of the data access operations on the cloud in a trusted mode.

Few of the schemes presented in the classification execute the entire security operations on the mobile device. The rest of the schemes delegate the security management operations on the cloud, trusted entity under the control of client organization, or trusted entity ender the control of the third party. In the category of local execution, a small number of schemes focus on the reduction of computational complexity of cryptography algorithms for reducing the processing burden from the mobile device.

Zhao et al. [17] proposed a method for trusted data sharing on untrused cloud servers using Progressive Elliptic Curve Encryption(PECE) scheme. The PECE allows the encryption of message multiple times with different keys that can be decrypted in a single run with a single key. Ren et al. [20] provide the security features for mobile user in the cloud environment using the traditional cryptography methods. The focus of the proposed schemes is to reduce the computational complexity of the cryptography operations instead of offloading the computationally intensive operations on the cloud. The size of the ciphertext grows linearly with increase in number of ciphertext attributes. [26] that involves more pairing evaluation exponential operations while decrypting the ciphertext. However, the data owner has to transform the plaintext into ciphertext, and vice versa that involves execution of computationally intensive multiplication and exponential operations of large numbers on resouce constraion mobile device.

[27-31] propose proxy re-encryption algorithm to confirm the security of the data in the cloud, which can alleviate the client's burden, and enhance the confidentiality of cloud data. However, there are two major disadvantages with these techniques. First, for re-encryotion, the data owner must obtain user's public key before uploading. Second, because the same plaintext is used with different keys generated by proxy, therefore, the storage overhead becomes excessive.

A manager-based re-encryption scheme (MReS) defined in [16] achieves the data confidentiality for the mobile users by using the proxy re-encryption strategy. The proxy re-encryption helps the mobile user to delegate the data access operation on third party. Another cryptographic scheme defined by Tysowski and Hasan in [16] is cloud-based re-encryption scheme (CReS). The proposed scheme covers the limitations of the existing manager-based re-encryption scheme and the variations of manager-based re-encryption based on user-managed key ciphertext fetch by user. But these two schemes are more complex and need more time to re-encrypt. To optimize the cloud storage, safety transmission, minimize the cost and computational complexity, here we have proposed a new scheme as searchable re-encrypted data in SReCSM. This technique is more simple and faster when sorting the arrays using the most significant radix sort. This technique will reduce the cost of the data owners up to O(Nt*3) and the time complexity will be reduced up to the O (B), where B denotes the Bucket size of the data base.

III. CONSTRUCTION OF SRECSM MODEL

The transfer probability of Markov chain is introduced into the cloud storage. The stored procedure in the cloud system is similar to the Markov process. Therefore, a Searchable Re-encryption Cloud Storage Method based on Markov Chain is proposed in this paper, which can realize reliable and safe storage on mobile terminal.

From the probability point of view, the data is scheduled to be stored in the node with the lowest storage cost in real time. Firstly, the storage state of SReCSM is divided and three state models are obtained in the paper. According to the storage cost of the nodes for storing different size file, the state matrix in current time is got by using the steady state of Markov strategy. The state transfer probability matrix is calculated with the support of a sufficient number of samples. The storage state probability of SReCSM can be predicted quickly in the future. Therefore, the optimal

storage node is selected to implement cloud storage access on mobile terminal.

When a request is made by a mobile terminal, the storage node which has the lowest cost is selected for storing. The SReCSM can save time and improve storage efficiency effectively. And at the same time the load balancing of the system is also guaranteed.

A. The finite state space and the state distribution

The Markov chain is a discrete random process with Markov properties. The next state of the random process is only related to the current state, not its historical state. Therefore, the distribution of the future state of the random process depends only on the present, not the past. The SReCSM based on Markov chains consists of the following three parts $\{S, R, P\}$. In which S is the finite state space; R stands for the state distribution; P is the probability of state conversion.

The server in the cloud storage group is named storage node. Each node can store large files, medium files, and small files. Files stored over 100MB is called a large file, files no more than10MB are defined as small files, files between 10MB and 100MB are defined as secondary files. The storage nodes are different in hardware performance, load situation, network link state and so on. Therefore, each storage node has different storage cost for storing large data, secondary data, and small data. That is, the storage cost for storing different size of files in each node is different. In the cloud system, depending on the size of the storage data file, each node has different storage costs for large, secondary and small files, and then the optimal storage node is selected accordingly.

Based on the analysis above, the SReCSM based on Markov Chain is proposed in this paper. The processes of the storage that store three different sizes of data in a storage cluster are treated as space S which is discrete state. Moreover, in the procedures of data storing, the selection of the next storage node is only related to the storage cost of the current storage node. Therefore, the data stored procedure of cloud system conforms to the characteristics of Markov chain. In the SReCSM proposed in this paper, the state space S is expressed as the following three types.

State 1 (Small file status): In the state of small file status, the optimal storage node for storing small files in a storage group cluster is selected to implement data storage.

State 2 (Secondary file status): In the state of secondary file status, the optimal storage node for storing secondary files in a storage group cluster is selected to implement data storage. Secondary files are between a small file and a large file.

State 3 (Large file status): In the state of large file status, the optimal storage node for storing large files in a storage group cluster is selected to implement data storage.

The state space of the cloud storage based on Markov chain is defined in formula (1).

$$S = \{s_1, s_2, s_3\} \tag{1}$$

In formula (1), s_1 denote the storage cost for storing small file on each storage node in the cluster, s_2 denote the storage cost for storing secondary file on each storage node in the cluster, and s_3 denote the storage cost for storing large file storage on each storage node in the cluster.

Moreover, according to the probability rule, the more detailed the storage state of the cloud storage system is, the more explicit the guidance of the data storage solution prediction will be. However, the greater the burden of computing resources needed for partitioning the state of the cloud system storage scheme. As we can see, in implementing the choice of cloud storage solutions, the degree of partitioning the storage state in cloud system storage scheme is a compromise decision problem. The more detailed the storage state is the more simplification is needed to reduce the computational burden. And the more storage states are, the greater the redundancy of the storage scheme. Therefore, the approach of partitioning the storage state not only complicates the analysis, but also reduces

storage speed. In fact, storage status can't be considered completely, and it is difficult for the SReCSM to achieve multiple storage state ultimately. Therefore, in the study of the SReCSM, the partitioning of the storage state is appropriate. Obviously, the storage state defined in this paper can be divided into three state, those are large file status, middle file status and small file status. Whether for the actual data of the project or the simulation analysis, these three states are relatively simple and can better satisfy the requirement of data storage in cloud system.

The storage group has an irregular connection to the network. The distribution matrix of the storage state in the cloud system is defined as R. And each data in the matrix is the storage cost on a certain storage node for data storing. Therefore, according to the storage cost of the nodes in large, medium and small three different states, the distribution matrix $R(i)$ of the storage status at time t_i can be obtained. The matrix $R(i)$ is to be defined in formula (2).

$$R(i)=\begin{bmatrix} s_1(i) \\ s_2(i) \\ s_3(i) \end{bmatrix}^T = \begin{bmatrix} R[X(t_i)=s_1] \\ R[X(t_i)=s_2] \\ R[X(t_i)=s_3] \end{bmatrix}^T \tag{2}$$

B. The probability matrix of the state transfer

The number of states in the state space is denoted with n. Therefore, the transfer probability can be represented in the form of matrix, which is defined as the probability matrix of state transfer. The probability matrix P of the state transfer always keeps the same. Moreover, and P is the matrix of order n. If the interval among $t_1, t_2, \cdots, t_n$ are always the same, according to homogeneity, P is defined in (3).

$$P=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \tag{3}$$

Obviously, the transfer probability matrix has two properties, which are shown in formula (4).

$$\begin{cases} p_{ij}(\Delta t)\geq 0 & i,j=1,2,\cdots,n \\ \sum_{j=1}^{n} p_{ij}(\Delta t)=1 & i=1,2,\cdots,n \end{cases} \tag{4}$$

In equation (4), p_{ij} refers to the probability of transition from the state s_i to the other state s_j. The sum of each row in the probability matrix is one. Moreover, the sum of the probabilities from the storage state s_i to all the other storage states s_j is one too. According to the analysis previous, each storage node in a storage group has different storage costs for storing three different size files. Therefore, in the cloud system, if a large, medium, and small files are need to be stored in the storage node, the higher the cost of integrated storage, the less likely this node is used to store data. On the contrary, the smaller the storage cost of the node for storing large, medium and small files, the more likely the node will be selected. The state space of SReCSM is represented in (1). The storage status transfer can be shown in figure 1.

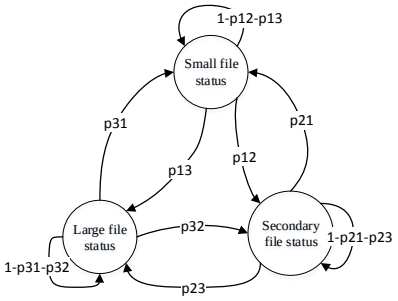


Figure 1. Storage status transition diagram

According to the probability values of the transformation in three different states of the storage node, the probability matrix of the state transfer in the SReCSM is shown in formula (5).

$$P=\begin{bmatrix} 1-p_{12}-p_{13} & p_{12} & p_{13} \\ p_{21} & 1-p_{21}-p_{23} & p_{23} \\ p_{31} & p_{32} & 1-p_{31}-p_{32} \end{bmatrix} \quad (5)$$

C. Algorithm of SReCSM

Users access the network through a mobile terminal, and the mobile terminal will choose to communicate with the node which has the lowest cost in the storage group. Moreover, irregular connection is used in the storage group. Therefore, in order to improve the reliability of the cloud storage system, a multi-replica mechanism is introduced. The storage efficiency of the cloud system can be effectively improved by using Multi-replica mechanism. When a mobile terminal makes a request, a copy with the lowest cost can be chosen. However, in the traditional cloud storage system, the master copy is selected first only when the master copy is wrong. The request will then be sent to the backup copy regardless of the region. Therefore, this kind of storage will affect the speed of the cloud system. Based on this idea, the transfer probability of Markov chain is used in the SReCSM. The storage cost of each storage node is measured from the perspective of probability. The state matrix of one-step transfer will then be obtained by means of the probability matrix of the state transfer. Finally, the state matrix of multi-step transfer will be obtained to predict the probability of the storage cost at a certain time in the SReCSM. Therefore, the optimal storage node will be chosen to realize the data storage at the end.

The main idea of the SReCSM algorithm in the cloud system will be introduced in the following chapter. The states of the storage in the strategy are proposed for the implementation of the data storage. Moreover, in the SReCSM algorithm of the cloud system, the standard of storage status division is proposed, which is convenient for engineering implementation. At the same time, the transfer expression of the state corresponding to the storage state is obtained. Then, on the base of the historical data stored in the cloud system, the probability matrix of state transfer is obtained. The probability and reliability of the storage status in the cloud system at some time in the future are predicted by using the probability matrix of the state transfer and the matrix of the storage state at the moment.

The steps of the algorithm of the SReCSM based on Markov chain are described as follows:

Step 1

The distribution matrix $R(0)$ of the SReCSM at the initial time t_0 is obtained.

Step 2

According to the distribution matrix of the storage state and the probability matrix P of the state transition at t_0 time, the state distribution matrix of one-step transfer $R(1)$ at next time will be got after Δt time. The state distribution matrix of one-step transfer $R(1)$ is shown in formula (6).

$$R(1)=R(0)P \quad (6)$$

Step 3

Obviously, the storage distribution of the cloud storage system after Δt time is shown in formula (7).

$$R(2)=R(1)P=R(0)P^2 \quad (7)$$

After a period of time, about m times Δt time, the storage status of cloud system is shown in formula (8).

$$R(m)=R(m-1)P=R(0)P^m \quad (8)$$

Through the procedure above, $R(0)$ and P are known, the steady-state value of the cloud storage distribution in the system can be predicted quickly in the future after every Δt time interval.

According to the storage cost of data stored in different nodes, the storage cost matrix of the current time of the system is obtained, which is the state distribution matrix of the system. The state distribution matrix is shown in formula (9).

$$R(i)=\begin{bmatrix} s_1(i) \\ s_2(i) \\ s_3(i) \end{bmatrix}^T = \begin{bmatrix} R[X(t_i)=s_1] \\ R[X(t_i)=s_2] \\ R[X(t_i)=s_3] \end{bmatrix}^T \quad (9)$$

In formula (9), s_i represents the storage cost needed to select a storage node at time t_i. We can know from formula (9) that s_i is contained within the range of 0 to 1, which satisfies the inclusion relationship $s_i \in [0,1]$. And s_i is obtained by normalizing the correlation property, which is stored in the node at t_i time. There are three variables in related properties at time t. Three variables are represented by symbols q, d and l, which respectively represent the length, latency and packet loss of stored data respectively. The storage cost of the node will be got in formula (10).

$$s = \beta_q s_q + \beta_d s_d + \beta_l s_l \quad (10)$$

In formula (10), three variables q, d and l satisfies the relationship in formula (11).

$$\sum_{k=\{q,d,l\}} \beta_k = 1 \quad (11)$$

Because parameters including q, d and l are all cost indicators. Therefore, the smaller the value, the better the quality of the data stored in this node. The storage cost will be got in formula (12) in this way.

$$s_k = \frac{k_{\max} - k}{k_{\max} - k_{\min}} \left(k = \{q, d, l\}\right) \quad (12)$$

In formula (11), q is obtained by the length of the stored data block. The parameters d and l are obtained by timing first, and then calculated through the time. Start the timer when each data block is stored at the beginning of the storage till the response is received. Therefore, the evaluation of the storage quality in each storage node is carried out through the above method. Moreover, and the distribution matrix of the storage status is updated in stages. The pseudo-code of the algorithm in SReCSM is described in figure 2.

Algorithm 1 Cloude Storage Strategy

```
procedure CloudStorage(a, c)
    S = {1, 2, 3}                                  ▷ Initialize the state space
    R = {1, 2, 3, ..., N}                          ▷ Initialize the state distribution
    Tdsf = T0                                      ▷ Initialize storage time
    Tsto = T1                                      ▷ Initialize Markov time
    while (Tdsf > 0) do
        if ( ! IsEmpty ( NC ) ) then               ▷ Node cost is not zero
            {
              GetDataFromNode(&Data)
              GetCost(&m)                          ▷ Get cost of node m
              StoreData( m , Data)                 ▷ Store data in node m
            if (Node.Cost > 1) then
              {
                StoreDataToOtherNode(&Data.Cost − 1)
              }
            end if
            Refresh(NC)
            }
        end if
        Tdsf − = 1
    end while
    //Markov Phase
    Build TM(&P)                                   ▷ Build State transfer matrix
    Build RV(&R)                                   ▷ Build State distribution matrix
    while (! bQuit) do
        {
        //Getthebeststrategy Phase
        while (Tsto > 0) do
            if ( ! IsEmpty ( NC ) ) then           ▷ Node cost is not zero
                {
                GetCostFromNode(&Data)             ▷ Get next time cost of node
                GetNextStrategy(Policy, &d)        ▷ Get strategy
                StoreData( d , Data)
                if (Node.Cost > 1) then
                  {
                    StoreDataToOtherNode(&Data.Cost − 1)
                  }
                end if
                Refresh(NC)                        ▷ Update node cost
                }
            end if
            Tsto − = 1
        end while
        Tsto = T1
        Refresh TM(&P)                             ▷ Initialize the state transfer matrix
        Refresh RV(&R)                             ▷ Initialize the state transfer distribution
    end while                                      1
    return c
end procedure
```

Figure 2. Pseudo-code in SReCSM

Therefore, in the SReCSM based on Markov chains, the distribution of the state in the future will be got by means of the probability matrix P of the state transfer and the distribution matrix $R(0)$ of the state at the initial time. The distribution of the storage state at time t_n will be calculated and the choice of SReCSM will be realized through matrix operation.

IV. SEARCHABLE RE-ENCRYPTION METHOD

The central idea of the searchable re-encryption is to generate a re-encryption key and decrypted while the keyword matched. By using searchable re-encryption method, SReCSM has increased the storage requirement due to the storing of all encrypted data. The objective of the proposed searchable re-encryption method is to increase the storage requirement, flexibility and reduce the security issues, overhead ratio and minimize the cost and searching time. This technique will not provide the effective data utilization. This technique is more simple and faster when sorting the arrays using the most significant radix sort. This technique will reduce the cost of the data owners and the time complexity will be reduced. This technique is more efficient and requires more storage space for the data.

A. Symbol Definitions

Symbols in the Searchable Re-encryption method are defined in the in Table I.

TABLE I. SYMBOL DEFINITION

Symbol	Definition
PR(K)	Private key enabled by Data owners
PU(K)	Public key enabled by data owners
Kw	Keyword
Db	Database {all the data passed to the cloud servers will be stored here}
Ek	Editing Keyword
Ed	Encrypted Data
PR(K)Re	Re-encryptedPrivate key
PU(K) Re	Re-encrypted Public key
Ed Re	Re-encryptedEncrypted Data
Dd	Decrypted Data
Kw(pu)	Keyword assigned with the public key
Kw(pr)	Keyword assigned with the private key
Do	Data Owners (Sender)
Du	Data Users (Receiver)
Cs	Cloud Server (Third Party User)

B. Keyword Editing

By searching the keyword we can edit the keyword with the help of three notations. While editing the keyword it requires, insertion, deletion or substitution. Here assuming the notation as n and editing the keyword as Ek.

Case 1 If $Ek \sim= n(i)$

Inserting the character into the first place

For example:

Character string (old) = ‘‘BOK’’

If $i=3;\ Ek \sim= n(3)$

In third place have to insert new character

Character String (new) = ‘‘BOOK’’

Case 2 If $Ek < n(i)$

Deleting the character into the first place

For example:

Character string (old) = ‘‘BOK’’

If $i=3;\ Ek < n(3)$

In third place have to delete the letter

Character String (new) = ‘‘BO’’

Case 3 If $Ek > n(i)$

Substitute the character into the first place

For example:

Character string (old) = ‘‘BOK’’

If $i=3;\ Ek > n(3)$

In third place have to substitute a new character

Character String (new) = ‘‘BOK’’

These three different notations help in editing the keyword easily. If the data users apply the keyword with some mistakes, they can edit the keyword to recover their respective data by using these three different cases.

C. Searchable Re-encryotion

The central idea of the searchable re-encryption is to generate a re-encryption key and decrypte while the keyword matched. Re-encryption operates over two groups G_1 and G_2 of prime order q with a bilinear map e: $G_1 \times G_1 \to G_2$. The system parameters are random generators

$g \in G_1$ and $Z = e(g, g) \in G_2$. The Searchable Re-encryption can be defined in the following algorithms.

Algorithm 2 follows that the Searchable Re-Encrypted Keyword and it will search the keyword to decrypt the files. If the keyword matched, the data can be decrypted and can bereceived and accessed by the users.

Algorithm 2 Keyword Searching

Input: Keyword for $PR(K)$

Output: Keyword Matched

If $PR(K) = Kw$

Goto $Db(i)$

Else If $Ek > n(i)$ \\ case-3

Goto $Db(i)$

Else "Not Matching"

End If $Ek < n(i)$ \\ case-2

"Keyword Matched"

Initially private key contains the keyword and that keyword will be searched in the database. Keyword with the private key should be verified in the database of data cloud storage. Data owners will transmit the data to the cloud servers and the cloud server will store the data in the database. The data users will receive the data from the database using the keyword. If the keyword did not match, data user use the editing scheme for insertion, deletion and substitution. If the keyword matched, data users will receive the data from the cloud servers which is explained in the Algorithm 2.

Algorithm 3 Data sharing

Generate $PU(K)$ $PR(K)$

$PU(K) = \{Ed, Kw(pu)\}$

$PR(K) = \{Kw(pr)\}$

Do shares $PU(K)$ to Cs

Do shares $PR(K)$ to Du

Send Ed to Cs

Du sends $Kw(pr)$ to Cs

Cs verify $Kw(pr)$

If $Kw(pu)$ = $Kw(pr)$

Cs sends $\{Dd, Kw(pr)\}$ to Du

Algorithm 3 states that the data file sharing between the data owners and data users through the cloud server. Initially data owners will create the public key and private key. The public key will be shared to the cloud servers and the private key will be shared to the data users. Every private and public key has its own keyword based on that keyword the data users will retrieve the data. If the keyword of public key and keyword of private key is matched, cloud server will transmit the data to the data users.

Algorithm 4 Data encryption

Consider two prime numbers as x and y

Assign $z = x * y$, where z will be used for the modulo of private and public keys

Assign Euler's function as $E(z) = (x-1)(y-1)$

Consider an integer as i such that $1 < i < E(z)$ for all $\{i, E(z) = 1\}$

Where i and $E(z)$ are co-prime

Assign $D = \{f_1, f_2, \cdots, f_n\}$; f as files, D as data and n as number of files.

$D(z) = \{f_1(z), f_2(z), \cdots, f_n(z)\}$

Encrypted data,

$Ed = \{D(z) \, mod \, E_z; Kw, PU(K)\}$

Algorithm 4 states that the data encryption scheme and this follows the RSA algorithm to encrypt the data. Initially the module multiplication of the two prime numbers will be calculated and this states that the Euler's function and this integer value is co-prime. Each data will be encrypted with modulo of E (z). The encrypted data contains the public key and the private keyword and the each data will be modulo with the Euler's function.

Algorithm 5 Data Re-encryption

Generate $PR(K)_{Re}$, $PU(K)_{Re}$

Assign $D=\{f_1, f_2, \cdots, f_n\}$, $g \in G_1$; f as files, D as data and n as number of files.

$$D(z)=\{f_1(z), f_2(z), \cdots, f_n(z)\}$$

Re-Encrypted data,

$$Ed(C_A)_{\mathrm{Re}} = e(PU(K), \quad Ed(C_A))$$

Algorithm 5 states that the data re-encryption scheme and this follows the AES algorithm to encrypt the data. The mobile user 'A' generates the re-encryption keys for authorized user's list 'U' using users' public key and personal private key as shown below:

$$PU(K)_{\mathrm{Re}} = PR(K)_{\mathrm{Re}} = (g^{x_i})^{\frac{1}{x_A}}, \forall u_i \in U$$

The CSReSM generates the $r_I \in Z_q^*$ randomly and encrypts the message using the following procedure:

$$Z_{new}^{r_I} = \frac{Z^{r_I}}{PU(K)}$$

$$(Ed \cdot Z_{new}^{r_I}) = (Z^{r_I} \cdot M)$$

$$Ed(C) = Z^{r_I} \cdot M, \quad Ed(CA) = g^{r_I x_A}$$

The CSReSM uploads the encrypted message ' $Ed(C)$ ' and ' $Ed(C_A)$ ' on behalf of the mobile user 'A'. The CSReSM transforms ' $Ed(C_A)$ ' into ' $Ed(C_A)_{Re}$ ' using the re-encryption key $PU(K)_{\mathrm{Re}}$ and $Ed(C_A)$ as shown in the following equation:

$$Ed(C_A)_{\mathrm{Re}} = e(PU(K), \quad Ed(C_A))$$

$$= e(g^{\frac{x_i}{x_A}}, \quad g^{x_A r_I})$$

$$= e(g, \ g)^{x_i r_I}$$

Algorithm 6 Data decryption

Consider D key = $\{PR(K), Kw\}$

For $f = 1$ to N //Data files

$$Dd = \{Ed \ mod \ PR(K)\}$$

End for f

loop

Goto Dd

Algorithm 6 states that the data decryption algorithm to decrypt the data using the above steps. Initially data users uses the decryption key and that decryption key contains private key with keyword. Select the number of files and then decrypt the data using modulo function. Each data will be decrypted using the decryption key.

D. Security analysis

Assuming that the reliability of the cloud storage system is identified by symbol A. The time of encryption through different encryption algorithms is A_t, the encryption time A_t is reversed first, and after the normalization processing, A_j is got from A_t. According to the Markov chain, the storage cost of the different node is normalized to be the value A_k. The number of the storage states for the cloud storage is the value n the reliability model of the system is shown in formula (13).

$$A = [1-(1-A_j)^n][1-(1-A_k)^n] \tag{13}$$

It can be concluded from the analysis of the reliability model. When the valet tends to be infinite, which means $t \to \infty$, the storage cost of the node in the cloud system also tends to a certain stable value, and the state of storage strategy tends to be stable. When the value of A_j and A_k are more closer to 1, and the value of n is more large, the cloud storage system will be more reliable and with higher security.

V. PROTOTYPE SYSTEM AND EXPERIMENT

A. Prototype system

HDFS and Dynamo are reliable solutions that are commonly used in the cloud storage system. HDFS is a distributed file system that is suitable for running on common hardware. Moreover, HDFS has good fault tolerance and can be used for inexpensive hardware. The program in HDFS has a lot of data sets. The file size in the HDFS is typically gigabyte to terabyte. As a result, terabytes of large files can be supported in HDFS through higher aggregated data bandwidth. Therefore, hundreds of nodal devices can be contained in a cluster, which allowing the terabytes of large files to be supported in it. The consistency hash algorithm is used by Dynamo. At that time, it's not the exact hash value, but a range of hash values. When the hash value of the key is in this range, it will be searched clockwise along the loop, and the first node encountered is what we need. The consistency hash algorithm is improved by Dynamo, and in the ring，a set of devices are acted as a node rather than only one device is acted as a node. The synchronization mechanism is used to achieve the consistency of the data.

In HDFS, numbers of the copies are set to be three. Whether the data would be stored in the node or not depends on the capacity of the node. The greater the capacity of the node is, the greater the probability that the data will be stored in this node. Therefore, when the capacity of the node is very different, the node with large capacity in the system would be overloaded. According to the analysis above, the storage cost of each storage node in the cloud system is measured from the perspective of probability. Moreover, the dynamic copy mechanism is adopted to adjust the number of copies in the storage strategy and their placement in real time according to the system performance requirement, load and so on. With the probability of passing the Markov chain, the dynamic replica mechanism can be used to complete data storage in time. The reliability and availability of SReCSM are improved effectively.

The storage group is used as the hardware architecture of the cloud storage system. Moreover, the storage group system is connected in the irregular network. The architecture of the storage group is shown in figure 3. The PC is used as a storage medium in the storage group. However, the reliability of the PC is not high, and it will even fail when the data are stored. Therefore, a copy is required to ensure that the data is reliable. According to the reliability criteria of the system, the storage state of the cloud system is divided into three types: small file status, large file status and secondary file status between the two.

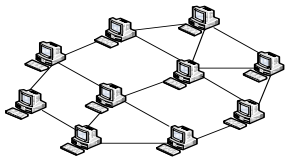


Figure 3. The system architecture diagram of the storage group

In the storage group system, all information about adjacent nodes is stored in every PC. The storage nodes needed can be found quickly through querying the information stored in the nodes. According to the transfer probability matrix and state distribution matrix of the SReCSM described precious, the storage strategy of a certain time is predicted.

The structure of storage space in the storage group is ring, and at the same time, the method of the unified addressing is adopted. In the storage group, the difference in performance of the PC can be offset by the virtual contiguous storage space. First, the hash algorithm message-digest is used to implement system address conversion. The actual physical address is processed and converted to 32-bit information string through the MD5 algorithm. And then these information strings are stored in the virtual continuous

address. Thus, the differences in performance between devices will be offset.

The converted address is mapped to the virtual storage space loop of the storage group through the MD5 algorithm. The device is found in the clockwise direction, and then the data is stored in the first PC mapped. Therefore, the data is backed up to two adjacent PC. The larger the amount of data in the system, the more uniform the spatial distribution will be. The data are stored when the routing of the corresponding PC and adjacent PC are updated. The routing information table is shown in Table II.

TABLE II. ROUTING TABLE

Field	Type	Length	Note
ID	int		Serial number
fname	varchar	255	Filename
fsize	int		File size
IP	varchar	15	IP address

The IP address of the PC device where the file replica located is stored in the IP field in the routing information table. The IP field is the routing information for the adjacent PC. However, once a node fails, all the information stored in the node are backed up and the routing information of the adjacent node is modified in time. According to the principle of the consistency hash algorithm, the storage space of the new PC device will be mapped to the new virtual address space when a new device needs to be added to the storage group. The existing space on the ring will not be changed, and this method can be very effective in avoiding the vibration of the address space. Meanwhile, the routing information on the adjacent PC is updated. The process of adding a PC is as shown in figure 4.

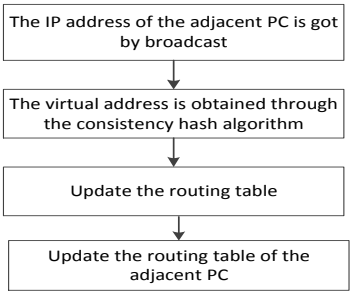


Figure 4. The process of adding a PC

B. Experiments and result analysis

The proposed scheme SReCSM Searchable Re-encryption Cloud Storage Method based on Markov Chain was developed by using the Java coding in the Linux platform. Finally evaluation of the overall performance of the proposed scheme in the real time data applications is done. This proposed scheme involved data owners and data users. Data owners create the encrypted data and then re-encrypted data will be forwarded to the cloud storage servers. Data users will decrypt the data using the keyword and private key.

In this section, we analyzed the prediction, the searching time, searching efficiency and storage space while performing moving, copying, encryption, decryption, and re-encryption operations. The final results are compared with the existing techniques such as CReS Cloud-based Re-encryption Scheme, MReS Searchable Encrypted Data File Sharing scheme.

1) Prediction of SReCSM

SReCSM based on Markov chain is processed by class Dynamo system model. On the basis of the thought and model mentioned above，the algorithm is implemented in python and verified by example. The simulation experiment is only to verify the Markov characteristics of distributed storage strategy.

At first, the Monte Carlo method is used to simulate 8000 times and 300 intervals each time. The distribution of storage state in these intervals and the change of storage status between adjacent intervals are counted. The state transfer probability matrix of the stored strategy is shown in formula (14).

$$P = \begin{bmatrix} 0.986 & 0.013 & 0.001 \\ 0.008 & 0.986 & 0.006 \\ 0.002 & 0.185 & 0.813 \end{bmatrix} \tag{14}$$

The meaning of the state transfer probability matrix P is described as follows. Suppose the last time the cloud system is in a small file storage state. At the next moment, the probability of 0.986 is kept in a small storage state; the probability of 0.013 is transferred to the secondary storage

state; and the probability of 0.001 is transferred to the large storage state.

The storage group in figure 2 contains 10 nodes. Because the hardware performance, load situation and network link state of the storage nodes in the cloud system are different, the storage cost of each node is different. Assuming at the initial moment t_0, the storage costs for large, medium, and small storage states are s_1, s_2 and s_3 respectively. The storage status distribution matrix $R(0)$ of the cloud storage system at time t_0 can be obtained through formula (2). The storage status distribution matrix $R(0)$ is shown in formula (15).

$$R(0)=\begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}^T = \begin{bmatrix} 0.10 & 0.20 & 0.70 \\ 0.35 & 0.20 & 0.45 \\ 0.15 & 0.45 & 0.40 \\ 0.20 & 0.30 & 0.50 \\ 0.35 & 0.30 & 0.35 \\ 0.25 & 0.25 & 0.50 \\ 0.30 & 0.35 & 0.35 \\ 0.10 & 0.15 & 0.75 \\ 0.20 & 0.35 & 0.45 \\ 0.25 & 0.30 & 0.45 \end{bmatrix} \quad (15)$$

The formula (10) (11) is substituted into formula (6), and a one-step storage state distribution matrix is obtained, which is shown in formula (16).

$$R(1)=\begin{bmatrix} 0.10160 & 0.32800 & 0.57040 \\ 0.34760 & 0.28500 & 0.36740 \\ 0.15230 & 0.51965 & 0.32805 \\ 0.20060 & 0.39090 & 0.40850 \\ 0.34820 & 0.36510 & 0.28670 \\ 0.24950 & 0.34225 & 0.40825 \\ 0.29930 & 0.41370 & 0.28695 \\ 0.10130 & 0.28795 & 0.61075 \\ 0.20090 & 0.43095 & 0.36815 \\ 0.24980 & 0.38230 & 0.36790 \end{bmatrix} \quad (16)$$

After a multi-step transition, the storage state distribution matrix $R(i)$ at time t_i is predicted quickly, which is shown in formula (17).

$$R(i)=\begin{bmatrix} 0.1139180 & 0.6278466 & 0.2582362 \\ 0.3425064 & 0.4889122 & 0.1685814 \\ 0.1644820 & 0.6808290 & 0.1546892 \\ 0.2072915 & 0.6045806 & 0.1881279 \\ 0.3444450 & 0.5212299 & 0.1343350 \\ 0.2520400 & 0.5606398 & 0.1873201 \\ 0.2996965 & 0.5651607 & 0.1351428 \\ 0.1129479 & 0.6116927 & 0.2753594 \\ 0.2082608 & 0.6207354 & 0.1710047 \\ 0.2530094 & 0.5767937 & 0.1701969 \end{bmatrix} \quad (17)$$

The node with the lowest storage cost can be quickly selected by using the storage status distribution matrix introduced in equation (17), which can realize real-time data storage.

2) Uploading and Downloading of SReCSM

The data response tests are performed on file upload, file copy and file movement, for large files and small files respectively. Experimental results demonstrate that the page is properly displayed, and the response time of the login page is basically completed within two seconds. The percentage of the response time for the transaction is shown in figure 5.

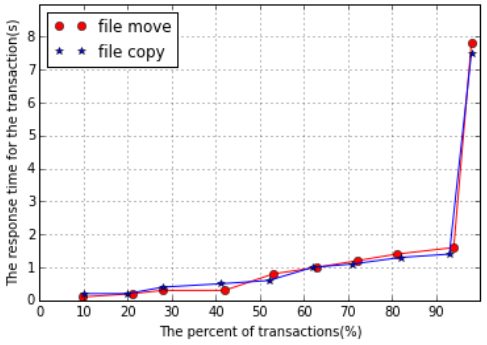


Figure 5. The percentage of the response time for the transaction

Through the analysis of figure 6, it can be known that 94 percent of transactions in a mobile cloud storage system can be implemented quickly within two seconds. The experimental results show that the system responds fast. The

average response time of the transaction is obtained from the diagram.

Although one of the response timing of transaction is longer, however the response time for most other transactions is acceptable. When this happens, it is thought that the performance of the mobile cloud storage system is better. Table III and Table IV are the test result.

TABLE III. THE PERFORMANCE OF THE MOBILE END UPLOAD THE DATA

type of test	utilization rate of Mobile CPU（%）	transmission rate （Mbps）	utilization rate of Mobile /transmission rate
Raw data	16.436	3.57020	4.603663
After the encryption	38.432	2.36015	16.283711

TABLE IV. THE PERFORMANCE OF THE MOBILE END DOWNLOAD THE DATA

type of test	utilization rate of Mobile CPU （%）	transmission rate （Mbps）	utilization rate of Mobile /transmission rate
Raw data	14.681	3.90135	3.763056
After the encryption	35.221	2.76147	12.754439

The ratio of CPU occupancy to upload speed is shown in table III, which the data on the mobile side are tested before the encryption and after the encryption respectively. The ratio of CPU occupancy to download speed is shown in table IV, which the data on the mobile side are tested before the decryption and after the decryption respectively. It can be known from the table III and the table IV, if the encryption and decryption mechanism are used for HDFS transmission, then the CPU utilization will be increased by an average of 22% ~ 25% and the overall file transfer rate will be reduced by 30% ~ 35%. As we can see，when the encryption and the decryption mechanism are used, more than three times the performance loss can be caused on the mobile end side.

3) Encryption and Re-Encryption

After applying the keywords, based on that proposed scheme it will search the keyword. If it matches, data users will receive their respective data. Based on the searching time and the storage space comparison graph has discussed below. Data users considered for the proposed scheme is 500 users and the available search keyword vary from 500 to 5000. Cloud servers storage space obtain more than 100 bytes and it uses the database to store all the encrypted data and the keywords.

There are two questions to be considered: One is the impact of encryption and decryption on file speed. The other is the impact of encryption and decryption on the performance of the client host. The experimental data are listed in Table V, which includes the time spent on encrypting the different sizes or different type files by using SReCSM and the time spent on transmitting the file in HDFS.

TABLE V. TIME COMPARISON ON ENCRYPTION AND DECRYPTION BY USING SRECSM

File size （M）	File type	SReCSM encryption （ms）	HDS upload （ms）	SReCSMdecryption（ms）	HDS download （ms）
3.07	pdf	1050	2685	370	2800
3.22	MP3	1178	2600	478	2830
23.8	mkv	3238	5930	2648	6290
25.8	doc	3140	5260	2163	6400
166.518	rmvb	23830	46400	16500	42460

It can be concluded from the above test data, the time spent on encryption or decryption by using SReCSM is regardless of the file type.

The time comparison on SReCSM encryption and re-encryption is shown in figure 6. We can see from figure 6 that there is little difference between encryption time and re-encryption time.

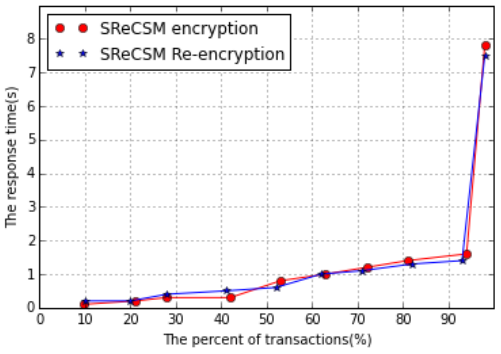


Figure 6. Time comparison on SReCSM encryption and Re-encryption

The same size files are encrypted by different CReS Cloud-based Re-encryption Scheme, MReS Manager-based Re-encryption Scheme, or SReCSM algorithms and then the re-encryption time is different, as shown in table VI and figure 7. The 167.58 MB file in table VI is the test case.

TABLE VI. TIME COMPARISON FOR DIFFERENT ALGORITHM ENCRYPTION

File size（M）	MReS Re-encrypt（ms）	CReSRe-encry pt（ms）	SReCSMRe-en crypt（ms）
3.04	1048	770	720
23.15	3230	2901	2600
80.35	12010	10230	8560
167.58	23820	23612	23598

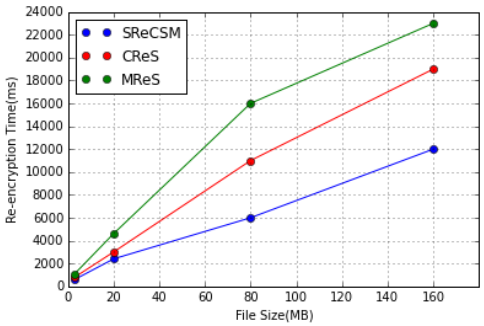


Figure 7. Comparison of Re-encryption time

In the SReCSM proposed in this paper，the time of encryption or decryption is relatively short. File transferring have little impact on total time loss and user experience. It may take a relatively long time to encrypt files by using the CReS, which cause a significant additional time overhead for HDFS. However, the encryption time that MReS encrypting the file was not significantly increased compared to SReCSM. Besides the impact on overall transmission rates, the impact of encryption and decryption on mobile performance is also important.

In the next experiment, we compared the searching time, searching efficiency and storage space while performing the encryption and re-encryption operations.

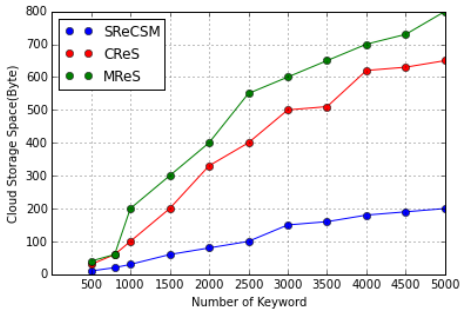


Figure 8. Storage spaceversus number of keyword

Figure 8 shows that the comparison graph of storage space in different algorithms. When there is increasing of the number of keywords, it requires more storage space. The existing algorithms CReS, MReS have require more storage space. But the proposed Searchable Re-encryption Cloud Storage Method (SReCSM) reduce the storage space requirement and utilize the data transfer effectively.

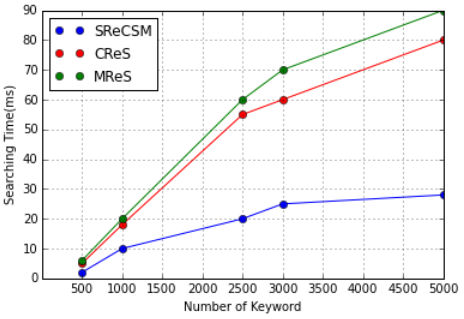


Figure 9. Searchingtimeversus number of keyword

Figure 9 shows that the comparison graph of the searching time and number of keywords. The number of keyword vary from 500 to 5000. When the number of keyword increases, searching time also increase. In previous techniques, CReS and MReS use the more searching time. However, SReCSM uses lesser time to search the data. If searching word is not matched, immediately the proposed SReCSM uses the editing values. Based on this different cases, the proposed SReCSM decreases the searching time.

MReS, CReS, and SReCSM offload the re-encryption operations on cloud. Therefore, in this experiment we examined the turnaround time and energy consumption on cloud while performing the re-encryption operations. The experimental results are shown in Fig. 10.

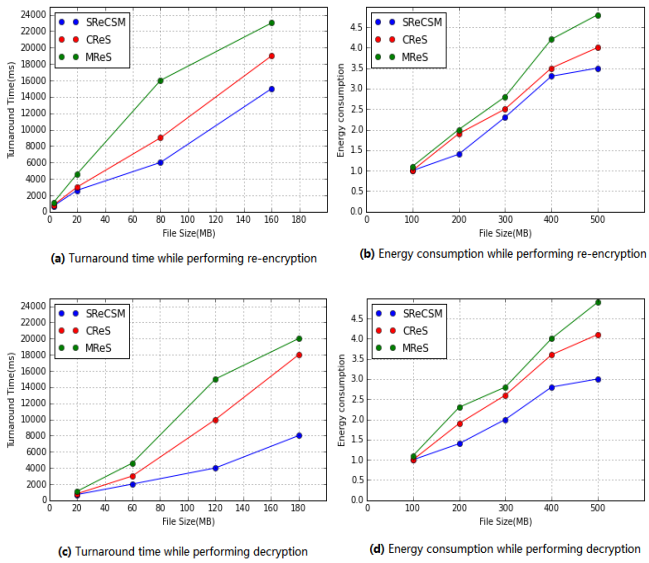


Figure 10. Comparison while performing re-encryption and decryption

It can be observed from the results presented in Figs. 10a and b that the increase in the size of file increases the turnaround time and energy consumption for completing the re-encryption operations on the mobile device. The increase in turnaround time and energy consumption is due to the increase in number of re-encryption operations while increasing the number of files. However, Figs. 10c and d show that the increase in the size of file increases the turnaround time and energy consumption for completing the decryption operations on the mobile device. The increase in turnaround time and energy consumption is due to the increase in number of decryption operations while increasing the size of files.

Using the reliability model formula (13) of the cloud storage system proposed in 4.4, combine the time required for processing the same size of file in table IV, when a different algorithm CReS, MReS and SReCSM is used, the encryption time required for encrypting file, after that the encryption time is reversed, A_j then be got after the encryption time is normalized. In the same way, after normalizing, storage cost A_k is got. If both A_j and A_k are closer to 1, and the number of storage state in the cloud storage system is larger, then the reliability of the system is higher. According to the above analysis，the data in one hour is sampled continuously, combined with the data in table IV and table V, the reliability contrast diagram for SReCSM is shown in figure 11.

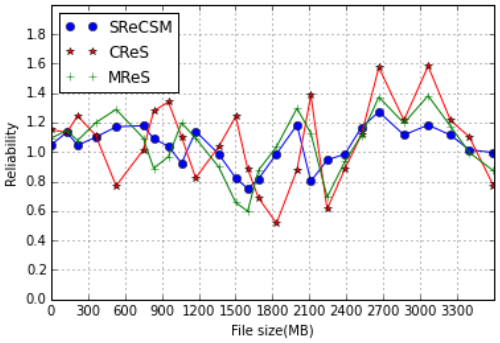


Figure 11. The reliability contrast diagram

It can be know the reliability of the system through different algorithm by comparing the data in figure 11. By using the CReS re-encryption， the reliability values are almost maintained at one. The reliability of the system is relatively high. That is， it has little impact on file transfer and user experience by using CReS re-encryption. It may take a relatively long time to re-encrypt files by using the MReS. And the reliability is very jitter. It is shown that the reliability is low with MReS re-encryption. However, the re-encryption time that MReS combined with CReS for re-encrypting the file was not significantly increased compared with CReS. The value of the reliability is consistent with the use of CReS, which can be maintained around one. It is concluded that the system is relatively reliable by using SReCSM encryption.

Through these simulation experiments, it is verified that SReCSM has a good user experience. It is also verified that the mechanism of SReCSM can effectively improve the efficiency of the cloud storage. When a mobile terminal makes a request, the optimal node is selected and then the time can be saved effectively.

In the SReCSM presented in this paper, the re-encryption and decryption has the following characteristics: transport security and storage security of the user data are guaranteed. The mobile finishes the re-encryption before calculating the checksum, so the re-encryption will not break the HDFS data integrity check mechanism. In the entire distributed file storage system, the re-encryption and decryption are

scattered to the various mobile devices. While this will cause some performance damage to the mobile, there is no additional performance penalty for name node and data node.

VI. CONCLUSIONS AND FUTURE WORK

CReS and MReS re-encrypts the keyword to transmit safety. But these two schemes are more complex and need more time to re-encrypt. To optimize the cloud storage, safety transmission, minimize the cost and searching time, here we have proposed a new scheme as searchable re-encrypted data in SReCSM. This proposed searchable re-encryption method supports the periodical and secure prediction by using the steady Markov strategy in stages and determine the lowest storage cost. SReCSM increases the storage requirement, flexibility and reduce the security issues, overhead ratio and minimize the cost and searching time.

The SReCSM based on Markov chain proposed in the paper has high reliability proved through a series of simulation experiments. The comparison graph evaluate the turnaround time and energy consumption with the different size of files. By increasing the size of files, proposed SReCSM can achieve accurate predictions, reduce the storage space requirement and the re-encrypting time. And then the data is scheduled to be stored in the node with the lowest storage cost. Finally conclude that the SEDFS proposed in this paper has better security and reliability. And this SReCSM reduces the storage space requirement, security issues, searching time and increases the searching efficiency.

ACKNOWLEDGMENT

Foundation item: The Industrial research project of Science and Technology Department of Shaanxi Province(Grant No. 2016KTZDGY4-09); Laboratory fund of Xi 'an Technological University (GSYSJ2017007)

REFERENCES

[1] Karel, Ferreira, Denzil,Goncalves, Jorge,Kostakos, Vassilis, De Moor, Katrien: Mobile cloud storage: A contextual experienceVandenbroucke.In: MobileHCI 2014 - Proceedings of the 16th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services, p 101-110, September 23, 2014

[2] Mitsutaka Kimura,Xufeng Zhao, Toshio Nakagawa: Using Markov Renewal Processes.Principles of Performance and Reliability Modeling and Evaluation Reliability Analysis of a Cloud Computing System with Replication, pp. 401-423(2016)

[3] Choo, Kim-Kwang Raymond: Mobile cloud storage users. In: IEEE Cloud Computing,v 1, n 3, p 20-23, September 1, 2014

[4] Iliadis, I., Sotnikov, D., Ta-Shma, P., Venkatesan, V.: Reliability of geo-replicated Cloud storage systems. In: 2014 IEEE 20th Pacific Rim International Symposium on Dependable Computing, pp. 169–179 (2014)

[5] Jeyanthy, C.,Shaji, R.S.,Jayan, J.P., Symmetric key based cryptic scheme for mobile cloud storage.In: Global Conference on Communication Technologies, GCCT 2015, p 571-575, November 30, 2015

[6] Jung, Kye-Dong, Moon, Seok-Jae, Kim, Jin-Mook: Data access control method for multimedia content data sharing and security based onXMDR-DAI in mobile cloud storage.Multimedia Tools and Applications, v 76, n 19, p 19983-19999, October 1, 2017

[7] Chekam, T.T., Zhai, E., Li, Z., Cui, Y., Ren, K.: On the synchronization bottleneck of OpenStack Swift-like cloud storage systems. In: IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications, pp. 1–9 (2016)

[8] Li, L., Li, D., Su, Z., Jin, L., Huang, G.: Performance analysis and framework optimization of open source cloud storage system. China Commun. 13(6), 110–122 (2016)

[9] Iliadis, I., Sotnikov, D., Ta-Shma, P., Venkatesan, V.: Reliability of geo-replicated Cloud storage systems. In: 2014 IEEE 20th Pacific Rim International Symposium on Dependable Computing, pp. 169–179 (2014)

[10] Yu, Xiaojun, Wen, Qiaoyan: Design of security solution to mobile cloud storage. Advances in Intelligent and Soft Computing, v 135, p 255-263, 2012

[11] Han, Lin; Huang, Hao; Xie, Chang-Sheng: Multi-path data prefetching in mobile cloud storage. In: Proceedings - 2014 International Conference on Cloud Computing and Big Data, CCBD 2014, p 16-19, March17, 2014;

[12] Lee, Giwon; Ko, Haneul; Pack, Sangheon: An Efficient Delta Synchronization Algorithm for Mobile Cloud Storage Applications. IEEE Transactions on Services Computing, v 10, n 3, p 341-351, May-June 2017;

[13] SystemWang, Yan; Wang, Jinkuan: An Optimized Replica Distribution Method in Cloud Storage. Journal of Control Science and Engineering, v 2017

[14] Zhang, Rui; Lin, Chuang; Meng, Kun; Zhu, Lin: A modeling reliability analysis technique for cloud storage system. In: International Conference on Communication Technology Proceedings, ICCT, p 32-36, 2013, ICCT 2013 -Proceedings of 2013 15th IEEE International Conference on Communication Technology

[15] Chen, Ming-Hung; Tung, Yu-Chih; Hung, Shih-Hao; Lin, Kate Ching-Ju; Chou, Cheng-Fu: Availability Is Not Enough: Minimizing Joint Response Time in Peer-Assisted CloudStorage Systems. IEEE Systems Journal, v 10, n 4, p 1424-1434, December 2016

[16] Tysowski, P.K., Hasan, M.A.: Re-encryption-based keymanagement towards secure and scalable mobile applica-tions in clouds. IACR Cryptology ePrint Archive 668, 2011(2011)

[17] Zhao, G., Rong, C., Li, J., Zhang, F., Tang, Y.: Trusteddata sharing over untrusted cloud storage providers, pre-sented at the IEEE Second International Conference onCloud Computing Technology and Science (CloudCom’10), Washington, DC, USA (2010)

[18] Yang, J., Wang, H., Wang, J., Tan, C., Yu, D.: Provabledata possession of resource-constrained mobile devicesin cloud computing. Journal of Networks 6, 1033–1040(2011)

[19] Itani, W., Kayssi, A., Chehab, A.: Energy-efficient incre-mental integrity for securing storage in mobile cloudcomputing, presented at

the International Conference onEnergy Aware Computing (ICEAC ’10) Cairo, Egypt(2010)

[20] Ren, W., Yu, L., Gao, R., Xiong, F.: Lightweight andcompromise resilient storage outsourcing with distributedsecure accessibility in mobile cloud computing. TsinghuaScience & Technology 16, 520–528 (2011)

[21] Yu, S., Wang, C., Ren, K., Lou, W.: Achieving secure, scalable, and fine-grained data access control in cloud com-puting, presented atthe Proceedings IEEE(INFOCOM’10)NJ, USA (2010)

[22] Jia, W., Zhu, H., Cao, Z., Wei, L., Lin, X.: SDSM: Asecure data service mechanism in mobile cloud computing,presented at the IEEE Conference on Computer Commu-nications Workshops (INFOCOM ’11) Shanghai, China(2011)

[23] Zhou, Z., Huang, D.: Efficient and secure data storage operations for mobile cloud computing, presented at the 8thInternational Conference on Network and Service Management (CNSM ’12), AZ, USA (2012)

[24] Ateniese, G., Fu, K., Green, M., Hohenberger, S.: Improvedproxy re-encryption schemes with applications to securedistributed storage. ACM Trans. Inf. Syst. Secur. (TISSEC)9, 1–30 (2006)

[25] Zhang, Yuan; Xu, Chunxiang; Li, Hongwei; Liang, Xiaohui: Cryptographic Public Verification of Data Integrity for Cloud Storage Systems. IEEE Cloud Computing, v3, n5, p 44-52, 2016

[26] Emura, K., Miyaji, A., Nomura, A., Omote, K., Soshi, M.: Aciphertext-policy attribute-based encryption scheme withconstant ciphertext length. Inf. Secur. Practice Experience5451, 13–23 (2009)

[27] Purushothama, B.R; Shrinath, B.; Amberker, B.B. : Secure cloud storage service and limited proxy re-encryption for enforcing accesscontrol in public cloud. International Journal of Information and Communication Technology, v5, n2, p167-186, 2013

[28] Cui, Yihui; Peng, Zhiyong; Song, Wei; Li, Xiaojuan; Cheng, Fangquan; Ding, Luxiao: A time-based group key management algorithm based on proxy re-encryption for cloudstorage. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and LectureNotes in Bioinformatics), v8709 LNCS, p117-128, 2014

[29] Shao, Jun; Lu, Rongxing; Lin, Xiaodong; Liang, Kaitai: Secure bidirectional proxy re-encryption for cryptographic cloud storage. Pervasive and Mobile Computing, v28, p113-121, June 1, 2016

[30] Jiang, Linmei; Guo, Donghui: Dynamic Encrypted Data Sharing Scheme Based on Conditional Proxy Broadcast Re-Encryption for Cloud Storage. IEEE Access, v5, p13336-13345, July 13, 2017

[31] Wang, XuAn; Xhafa, Fatos; Hao, Wei; He, Wei: Non-transferable unidirectional proxy re-encryption scheme for secure social cloudstorage sharing. Proceedings - 2016 International Conference on Intelligent Networking and Collaborative Systems, IEEE INCoS 2016, p328-331, October 25, 2016

Design and Analysis of Thermoplastic Metal Detector RC Car with Wireless Charging

Haifa El-Sadi
Wentworth institute of technology
Mechanical engineering Department
Boston, MA
e-mail: elsadih@wit.edu

Derek Fernandes
Wentworth institute of technology
Mechanical engineering Department
Boston, MA

Matthew R. Cole, Aaron M. Denis
Wentworth institute of technology
Mechanical engineering Department
Boston, MA

Alberto Benhamu-Chocron
Wentworth institute of technology
Mechanical engineering Department
Boston, MA

***Abstract*—The RC Car industry is a growing industry that will always be a past time for the older generation. This research is focused on a specific type of RC Car which is a Metal Detecting one. Metal Detectors have been developed for many years. Through research there aren't many metal detecting RC Cars on the market. Currently it's extremely limited in range and depth of detection.**

The goal of this research is to improve, build and test a new RC-metal detector technology. A thermoplastic-ABS used to build the chassis of the RC- metal detector. ABS is easily machined, sanded, glued and painted. Finite element analysis is a powerful tool that allows us to quickly analyze and refine a design. When the Chassis fixed at front Differential (Displacement), the value of the maximum stress and the deflection are higher than when the Chassis Fixed at Rear Differential. The maximum deflection is shown as about 1.339 inches. This value is the resultant of the deflection in all three directions. This research includes, a cost analysis for each piece of the car, group goals for the RC Car and Metal Detector. A working prototype is designed before moving into the final design phase in order to assure that the best possible product can be produced. There is a wireless charging station that ease the process of recharging the battery.

Keywords-Metal Detector; Finite Element; Thermoplastic; RC

I. INTRODUCTION

The challenge for this project is to design and build a remote-controlled car that incorporates two features that no other car possesses. RC Car owners have tedious work to do every time the car needs to be charged. Some components must be taken apart to retrieve the battery to charge it. The wireless charging feature being implemented in the design will eliminate this extra work, while providing a smooth charging operation where no disassembling will take place.

The design and development of remotely operated solar-powered mobile metal detector robot is a rescue robot to autonomously operate in detecting the threat of land mines. During the First and Second World War, military forces deployed many bombs on land filed to fight between soldiers on the battlefield. There were many countries like Libya, Cambodia and Laos had explosive weapons that did not explode when fired or dropped on the ground. In fact, more than twenty thousand people have been killed or injured by unexploded bombs [1]. A remotely solar-powered mobile metal detector robot has been designed and implemented. The system is using RF communication with Atmega32 MCU in embedded system domain. The robot moves in particular direction using the handheld remote. The experimental work has been carried out carefully. The metal detector sensor worked as the required specification for the metal detection sensor. The testing demonstrated that the robot would not pose any performance problem for the installation of the metal detection robot such as the merits and drawbacks of mounting the sensor, cost, support vehicle, handling the cable between the robot and also easiness of the adjustment [2]. Nation et al. [3] demonstrated the accuracy of HHMD in the identification and localization of metallic foreign bodies. They proposed an emergency room foreign body protocol that uses HHMD as an early screening tool in triage in order to expedite the process of obtaining Otolaryngology consultation and potentially shorten the wait time to the operating room or discharge. In instances were outside films are previously performed, HHMD use may be able to minimize the overall radiation exposure to children by obviating the need for repeat radiographs. As the sensitivity is not 100%, a negative HHMD screening does not negate the need for a standard radiograph in order to avoid missed MFBs. HHMD is best suited for detection of coins, which accounts for the majority of the MFB ingestions, and may not be suitable for all metallic objects since the amount of metal may decrease its sensitivity.

Holm, Katja F et. al. [4] evaluated a commercially available metal detector for detecting CIEDs. Design. Observational study including pacemaker patients (n = 70)

and a control group without pacemaker (n = 95). The investigational device was a hand-held metal detector for detecting metal or electricity wiring. Results. The metal detector detected the pacemaker in all pacemaker patients and thus exhibited a sensitivity of 100%. The specificity of the metal detector was 86%, and the negative predictive value was 100%. Thirteen individuals without pacemakers were falsely identified as having an implanted device due to implanted prosthetic material or elements of clothing.

The objective of this research is to design a RC car that will provide a smooth wireless charging process and incorporate a metal detector. RC Cars require tedious work in order to charge the battery. It will offer a wireless charging station where no disassembling of the car will be required. Apart from technical innovations, an improved metal detector will be incorporated that will be useful on beaches and in parks to search for lost jewelry. A long-term goal with an unlimited budget is to use this car to detect mines for the army. This car could drive over the field before infantry and passenger vehicles drive over it to protect them from mines or IEDs.

II. Material and Cost

Acrylonitrile Butadiene Styrene (ABS) is an opaque thermoplastic and amorphous polymer. ABS becomes liquid at a certain temperature, 221 degrees Fahrenheit. They can be heated to their melting point, cooled, and re-heated again without significant degradation. Instead of burning, thermoplastics like ABS liquefy which allows them to be easily injection molded and then subsequently recycled. ABS is easily machined, sanded, glued and painted. This makes it a great material for prototyping. Table 1 shows the required material to build the RC-car metal detector.

TABLE I. THE MATERIAL AND THE COMPONENTS OF RC-CAR METAL DETECTOR

	Cost
Supplies	Battery - $180 Engine - $120 Suspension Kit - $80 Wheels & Tires - $60 Controller & Transmitter - $80 Receiver - $80 Metal Detector - $40 Wireless Charging Components - $100
Prototype	Pre- built RC Car - $80
Equipment	3D Printer - $0 (School Owned) CNC Mill - $0 (School Owned) Welder - $0 (School Owned) Soldering Iron - $0 (School Owned) General Tools - $0 (School Owned)
General Reserves	Possibility of broken components.

III. Finite Element Analysis of Material

SOLIDWORKS Simulation uses the displacement formulation of the finite element method to calculate component displacements, strains, and stresses under internal and external loads. The geometry under analysis is discretized using tetrahedral (3D), and solved by iterative solver. SOLIDWORKS Simulation using p adaptive element type, the solution has converged. The material parameters were obtained and the results were simulated. One of the most important inputs to the model is the elastic modulus E of the material. The elastic modulus defines the stiffness (resistance to deflection) of the material. Its value is determined from material tests. A material with a high value of E will deflect less than one with a lower value of E. By applying finite element analysis, we can accurately observe the stress distributions in the various layers of the material as shown in Figures 1, 2, 3 and 4. Figure 1 shows Chassis Fixed at Rear Differential, the highest stress is 4039 psi with mesh size was 0.2. However, the maximum deflection was 0.7 inches as shown in Figure 2. Simulates impact on underside of chassis.

ABS Plastic Material Data

Elastic Modulus: 290075.4753 psi, Shear Modulus: 46252.53454 psi, and Tensile Strength: 4351.13213 psi

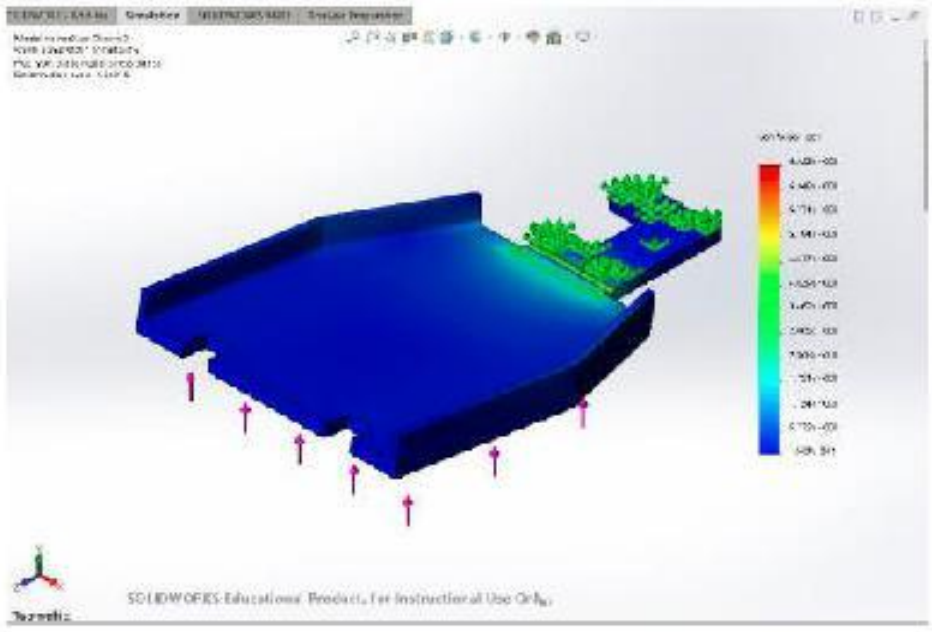


Figure 1. Chassis Fixed at Rear Differential (Von Mises)

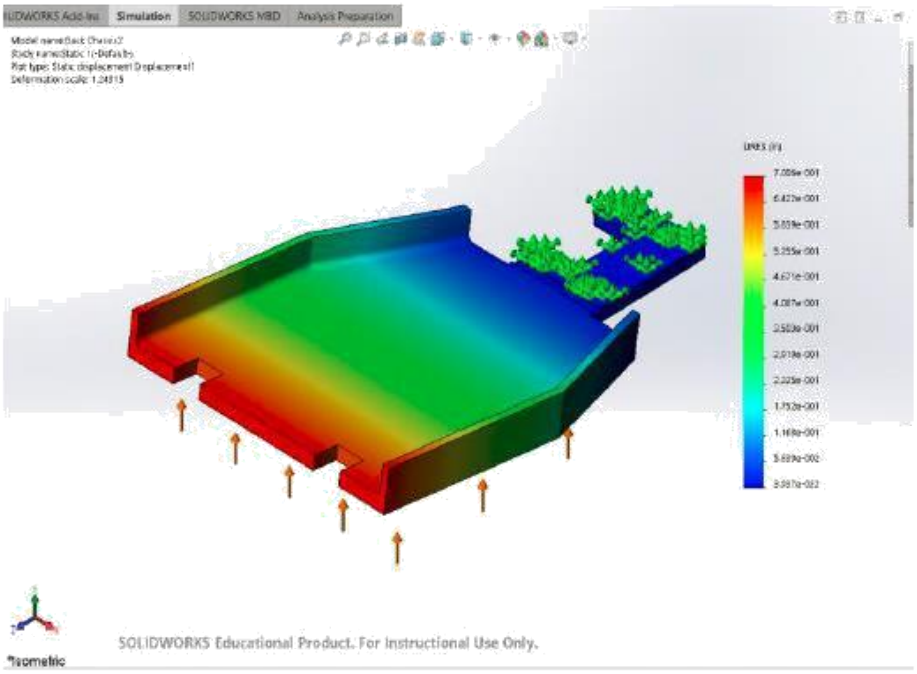


Figure 2. Chassis Fixed at Rear Differential (Displacement)

Figure 3 shows Chassis Fixed at front Differential (Displacement), the highest stress is 8375 psi with mesh size was 0.2. However, the maximum deflection was 1.339 inches as shown in Figure 4, it Simulates impact on underside of chassis

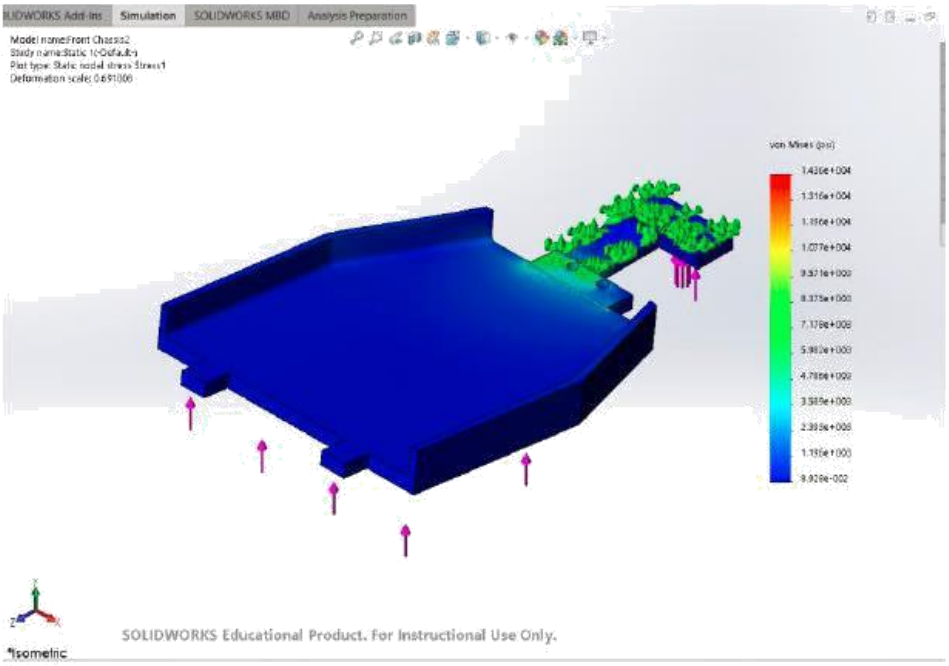


Figure 3. Chassis Fixed at Front Differential (Von Mises)

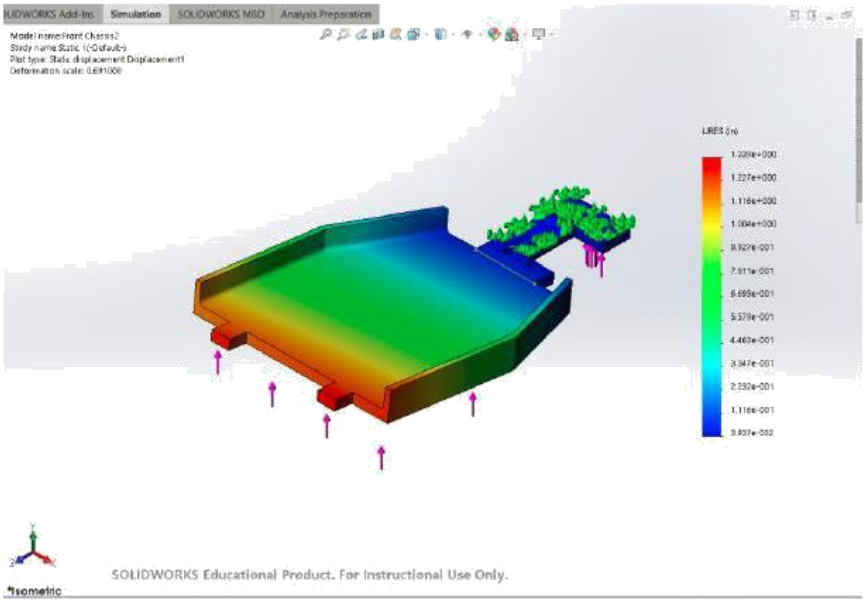


Figure 4. Chassis Fixed at front Differential (Displacement)

IV. DESIGN PROTOTYPE AND TESTING

One of the biggest challenges was designing a chassis from scratch that would house all the components necessary to operate a RC Car. The chassis also needed to be strong enough to not only hold the weight of the car and all its components, as shown in Figure 5, but also handle the torque and power that would be outputted from the car. The designed chassis was more than suitable to hold all the components and handle the force of the motor. There was more room than anticipated, which was useful for placing the components and keeping things away from all the moving parts. In the end, the RC Car is programmed and moves as planned. The metal detector detects deeper than expected. Finally, the wireless charging station works better than anticipated. Figure 7 shows the Wireless Charging Ramp Setup.

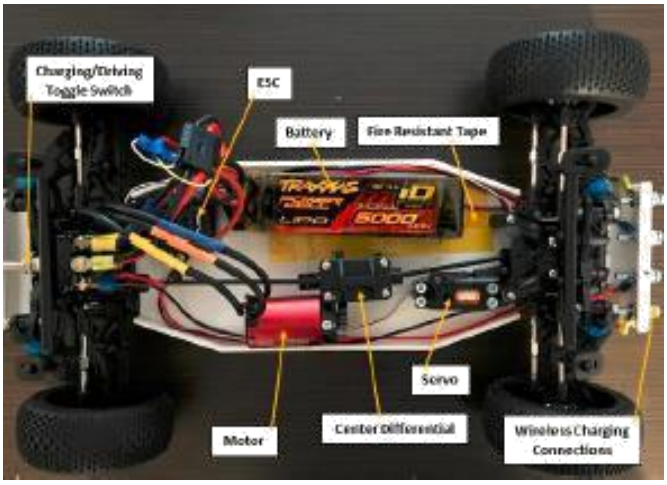


Figure 5. Detailed Picture of all Components

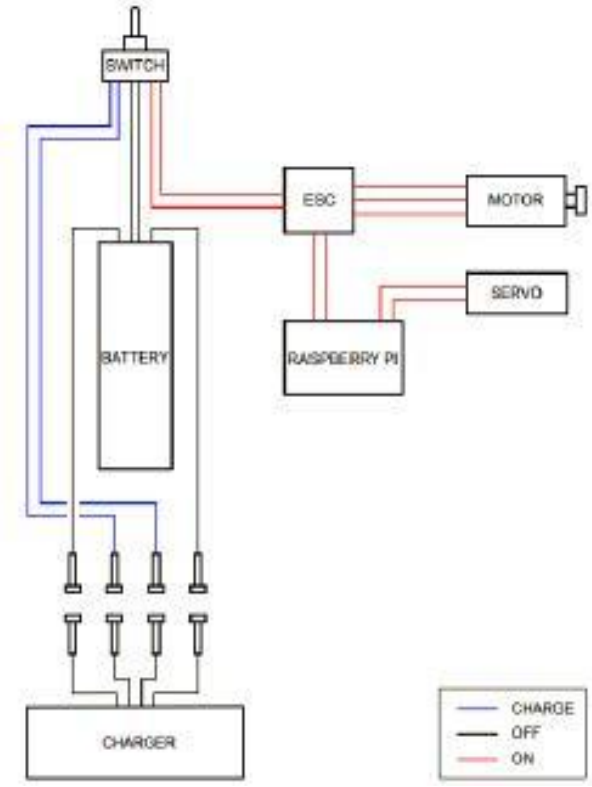


Figure 6. Layout of the Electrical Circuit

Figure 6 shows the layout of the electrical circuit of the RC-car metal detector. The main circuit is powered by a battery through a mechanical switch. If the switch is turned ON, power is supplied to a motor and a servo. The speed of the motor is regulated by an Electronic Speed Control Circuit (ESC). The servo on the hand is controlled by a Raspberry Pi computer board. The circuit also has four terminals that are used to connect to an external charger once the battery has depleted.

Figure 7. Wireless Charging Ramp Setup

Figure 8. Final prototype

V. CONCLUSION

This project aimed at creating a thermoplastic RC Car designed with a purpose, more than just an average recreational toy. This was bringing up the reality of a much greater idea. The wireless charging aspect of the RC Car will mostly be aimed towards small scale vehicles. However, the metal detection can be a very realistic and useful application for much bigger vehicles. The idea that an unmanned vehicle can be used for detecting not only metals, but harmful objects could potentially be a breakthrough for something like the military. IED's are a very serious issue for the military, being able to sniff these out without having to risk the lives of anyone would be huge. Completing this project with a limited budget would show a lot, especially when someone like the military could spend endless amounts of money in order to perfect the system for more dangerous and life like scenarios. FEA analysis was used to test the chassis, When the Chassis Fixed at front Differential (Displacement),

the value of the maximum stress and the deflection are higher than when the Chassis Fixed at Rear Differential.

VI. ACKNOWLEDGEMENTS

The authors sincerely appreciate the immense and highly valued support from the Dean of Engineering, Dean Fred Driscoll

REFERENCES

[1] Handicap International, Fatal Footprint: The Global Human Impact of Cluster Munitions. November 2006. Available from: URL: http://www.mineaction.org

[2] F.Y.C. Albert, C.H.S. Mason, C.K.J. Kiing, K.S. Ee, K.W. Chan, "Remotely Operated Solar-powered Mobile Metal Detector Robot", Procedia Computer Science, Volume 42, 2014, Pages 232-239

[3] Nation, Javan and Jiang, Wen, "The utility of a handheld metal detector in detection and localization of pediatric metallic foreign body ingestion", International Journal of Pediatric Otorhinolaryngology January 2017 92:1-6

[4] Holm, Katja F, Hjortshøj, Søren, Pehrson, Steen, Svendsen, Jesper Hastrup, Riahi, Sam, "Implanted cardiac devices are reliably detected by commercially available metal detectors", Scandinavian Cardiovascular Journal. Oct2013, Vol. 47 Issue 5, p271-274. 4p.

Traffic Identification and Traffic Analysis Based on Support Vector Machine

Wang Zhongsheng
School of Computer Science and Engineering
Xi'an Technological University
Xi'an 710021, Shaanxi, China
e-mail: wzhsh1681@163.com

Yang Sen
Shijiazhuang Campus, Army Engineering University
Shijiazhuang, Hebei, China
e-mail: yangsanmu@oec.mtn

Wang Jianguo
School of computer science and engineering
Xi'an Technological University
Xi'an 710021, Shaanxi, China
e-mail: wjg_xit@126.com

Gao Jiaqiong
Department of Computer Science
Sichuan Vocational and Technical College
Suining 629000, Sichuan, China
e-mail: 516719510@qq.com

***Abstract*—As the traffic generated by the increasing number of applications on the Internet is becoming more and more complex, how to improve the quality of service and security of the network is also increasingly important. This paper studies the application of Support Vector Machine(SVM) in traffic identification to classify network traffic. Through data collection and feature generation methods and network traffic feature screening methods, SVM is used as a classifier by using the generalization capability of SVM, and the parameters and kernel functions of the SVM are adjusted and selected based on cross comparison ideas and methods. Using the cross-validation method to make the most reasonable statistics for the classification and recognition accuracy of the adjusted support vector machine avoids the situation that the classification accuracy of the support vector machine is unstable or the statistics are inaccurate. Finally, a traffic classification and identification system based on SVM is realized. The final recognition rate of encrypted traffic is up to 99.31%, which overcomes the disadvantages of traditional traffic identification and achieves a fairly reliable accuracy.**

Keywords-Support Vector Machine (SVM); Traffic Classification; Feature Extraction; Kernel Function

I. INTRODUCTION

Due to the rapid development of the Internet, Internet business has greatly facilitated and enriched people's lives, learning and work, and has attracted more and more users. With the new application patterns (such as P2P) and application demand emerging in the Internet[1], the pressure of huge data transmission is becoming more and more heavy, and the occurrence of network failures is becoming more and more frequent, which leads to a series of network failures, such as packet loss, network congestion, and time delay in the process of data transmission. The maneuverability of the network is greatly reduced, the normal operation of the network is affected, and huge economic losses are incurred. Therefore, how to identify and classify the network traffic in real time helps the Internet service provider to understand the network operation status and optimize the network operation and management. It is of great significance.

The current popular network traffic identification technologies include traffic identification algorithms based on known ports [2]; traffic identification based on Deep Packet (DPI) [3-5]; traffic identification algorithms based on data flow behavior pattern [6]; traffic identification algorithms based on machine learning and so on.

The traffic classification method [7] has been widely proposed in the past few years. Initially, the type of data transmitted over the Internet is relatively small. The traffic identification technology is mainly based on port identification. That is, the general network protocol port number [8] is used to roughly classify traffic. For example, the protocol uses a fixed port. However, with the development of the Internet, merely relying on port identification technology has been insufficient to distinguish between more and more network applications and protocols. In 2004, an application layer load signature recognition method, the DPI technology [9], was proposed to extract the data message samples and determine whether the traffic belongs to the application by matching the signature of the unknown traffic. In recent years, the proportion of network traffic transmitted by encrypted text is increasing. DPI technology has been powerless for this part of the traffic. At present, the method of network traffic recognition based on machine learning [9] shows a higher accuracy.

Machine learning is an important tool for the study of network traffic identification. Dong S and others described the current popular machine learning method [10]. After comparing and evaluating the clustering algorithm [11], it was found that the feature selection algorithm [12] was better for supervised machine learning[13,14]. DBSCAN algorithm [15] of unsupervised clustering algorithm has higher precision.

Since the development of a complete classification architecture [16] for real-time work on high-capacity links is limited, Este A[18] and others after demonstrating the computational time and the optimization steps required to handle different traffic traces, used machine learning techniques(SVM model[17]) to improve system performance and enable real-time traffic identification for high-speed networks. Zhao X proposed a P2P network traffic classification method based on support vector machine [19], using a statistical principle to divide the network traffic of four different types of P2P traffic applications (file sharing BitTorrent, media streaming PPLive, Internet phone Skype, instant messaging MSN), and studied network traffic statistics and SVM methods. The overall framework of P2P traffic classification based on SVM was introduced, how to obtain traffic samples and processing methods were described, and the traffic classifier was constructed, with an average accuracy of 92.38%.

Bernaille L and others divided the traffic classification mechanism into two phases [20]: offline learning and online classification. The offline learning stage uses the kMeans method [21,22,23] to divide the original traffic and give a description of each cluster and its application type; the online learning stage determines the application type of the new traffic according to the learning knowledge.

Ye M proposed a new method of identifying P2P traffic through data transmission behavior of P2P applications [24]. The data downloaded from the P2P host finds the shared data of the download stream and the online upload stream, and proposes a content-based partitioning scheme to divide the stream into data blocks.

Based on the above viewpoints and taking into account the excellent performance of machine learning and SVM in solving P2P traffic classification problems [25-29], this paper proposes a network traffic two classification method based on SVM. which is used to complete the network flow parameters obtained from the packet header after network traffic collection to classify Internet traffic into a wide range of application categories. In the selection process of feature vectors, it should be suitable for SVM algorithm and try to calculate independently of the protocol and port. Therefore, in this paper, we choose the number of packets, size characteristics, data flow time characteristics, flag bits and other information as a preliminary feature vector, through a plurality of classifier selection methods to obtain the optimized feature set. It is used to implement the initial identification of normal traffic in the network, reducing the workload of the feature value matching module, improving the efficiency of the network traffic identification system, and comparing with the method of identifying network traffic that only adopts the feature value matching. The experimental results using the traffic from the campus backbone network show that 99.31% accuracy can be achieved through regular biased training and test samples. When using bias-free training and test samples of the same

feature set, an accuracy of 96.12% can be achieved. In addition, since all feature parameters can be calculated from the packet header, the proposed method is also applicable to encrypted network traffic.

II. PROPOSED METHOD

A. *Support Vector Machine (SVM) model*

SVM is a machine learning method that is based on one of the statistical algorithms with good generalization ability. It is mainly used to solve small samples. The feature vectors of the data stream in the network are more or less, and too many features will affect the efficiency and accuracy of the SVM algorithm. Therefore, to reduce redundant features, feature combinations with high discrimination are selected as feature vectors. After completing the support vector machine network traffic classification identification code, statistics and evaluation of the operating efficiency and accuracy of the results are also required.

The identification of network traffic is essentially a pattern classification process and is mainly divided into the following three points:

1) Converting the actual problem into the high-dimensional feature space through the kernel function, so that in the high-level space, the hyperplane can be used to classify the data, and the classification decision function is constructed so that the nonlinear problem of the original dimension is converted into linear separable problem. The classification decision function is a linear combination of non-linear functions with support vectors as parameters. The classification function itself is only related to the number of support vectors, so the method of this kind of kernel function is very effective in dealing with the classification problem of high dimensional feature space.

2) Under the condition that the number of known training samples is small, the network traffic classification is converted into secondary optimization and improve the accuracy of classification. The initial threshold is determined by iterating feature subsets using the inter-class distances and intra-class distances of the features.

3) The optimization problem is coded by simulating the natural evolutionary process. The key point of coding is that the code must be able to represent all possible subsets of the feature set. The optimal hyperplane is used to optimize the learning ability of the classifier. This method does not need to rely on the prior probability of the network traffic samples and has better generalization.

When using SVM, classifiers with better generalization effects can be achieved by defining different kernel functions and relaxation factors. The optimization model is as follows:

Let the training sample set be: {（xi,yi）}，i=1，2，3，…n; map this sample set to the high-dimensional feature space and achieve regression, the following are obtained:

$$fx\omega T\ \emptyset x+b \tag{1}$$

(ω is the weight vector; b is the offset vector)

Convert equation (1) to the minimization problem. The objective function of SVM regression is:

minω2 +12Ci=1nδi2

s.t.

$$yi-\omega T\emptyset\, x+b=eii=1,2,3,\ldots n \tag{2}$$

In this formula, C is the penalty parameter; ei is the regression error. Through the Lagrangian operator, the corresponding dual problem is obtained as follows:

$$L\omega,b,\delta,\alpha=\min\text{[illegible]}|\omega|2+12$$
$$\gamma i=1n\delta i2+i=1n\alpha i(\omega T\emptyset\, x-b+ei-yi) \tag{3}$$

Set the kernel function Kxi,xj=∅(xi)T∅(xj), then use the nonlinear SVM regression model established by the RBF function. There are:

$$fx=i=1N\alpha iexp-xi-xj2\sigma 22+b \tag{4}$$

(σ is the width of the core)

B. *Finding support vectors in training samples*

Introduce the following rules to distinguish. Set the threshold of the support vector decision function λ=1 or λ=-1, Assume that the decision function in the detection process is fx=sgn{i=1n[aiKxi,x-sidi§]}, f (x)≠1 or f(x)≠－1, The x vector does not belong to the support vector or the x vector belongs to the support vector.

An initial support vector library trained from known flows. After the known flow rate is trained by the data acquisition module, the feature extraction module, the data preprocessing module, and the training module, a support vector is generated to perform feature analysis, and its characteristic word information is added to the support vector library. Various known P2P traffic passes through the above process eventually forms a multidimensional support vector group, and a known support vector library is also formed. Finally, the MSVM threshold is determined. If the threshold is equal to 1 (or -1), the detected network traffic is P2P traffic; otherwise, the detected network traffic is non-P2P traffic.

When selecting P2P traffic characteristics, the feature extraction should be able to reflect the difference of P2P traffic as much as possible. Different nodes in the network have different functions: Some nodes function as servers and provide resource transmission services to other nodes in the network. Some nodes function as clients and receive various services provided by the server. The nodes in the P2P network can serve as servers to other peer nodes, and can also serve as clients to receive services provided by other peer nodes. Therefore, node traffic with different functions and providing different services presents different behavior characteristics.

C. Support vector machine network traffic identification process

The network traffic identification based on vector machine is essentially making full use of the powerful capability of SVM to deal with non-linear multi-factor system to mine the internal rules and establish the complex non-linear relationship of network traffic change, so as to achieve accurate network traffic prediction. In the learning and classification process of the SVM model, the selection of kernel functions plays a decisive role in the training and classification performance. At present, several frequently studied kernel functions are: linear kernel, RBF(radical basis function) kernel and Gaussian kernel and so on. In this paper, RBF kernel is selected as the kernel function.

The overall strategy when selecting the kernel function and adjusting parameters is approximately the following steps: preparing a batch of classified data; splitting the data into two groups: a training group and a test group; using a training group to give a support vector machine for training and learning; The support vector machine predicts the classification of test group data and compares it with the actual classification of the number of test groups, calculates the classification accuracy, replaces the parameters, and then iterates again. If we do not use the cross comparison idea, it is very easy to cause the prediction result to be very good only in the case of a specific input. In other cases, the prediction of the parameter is not stable.

D. P2P traffic classification model based on SVM

Figure 1 shows the classification framework based on SVM in this paper. This paper firstly extracts and analyzes the traffic to extract several main characteristics of network traffic that are suitable for recognition in the support vector machine. Then, the data is preprocessed, and the known data set for the target problem is set as a training data set, and use an iterative process to train a classification model. The parameters of the model are continuously adjusted by a method of random optimization or analysis, so that it is closest to the actual situation of the training data set. After the model is trained, it can be used to identify unknown samples and dynamically adjust the training sample data by continuously searching for useful training samples to realize the entire network traffic identification based on SVM.

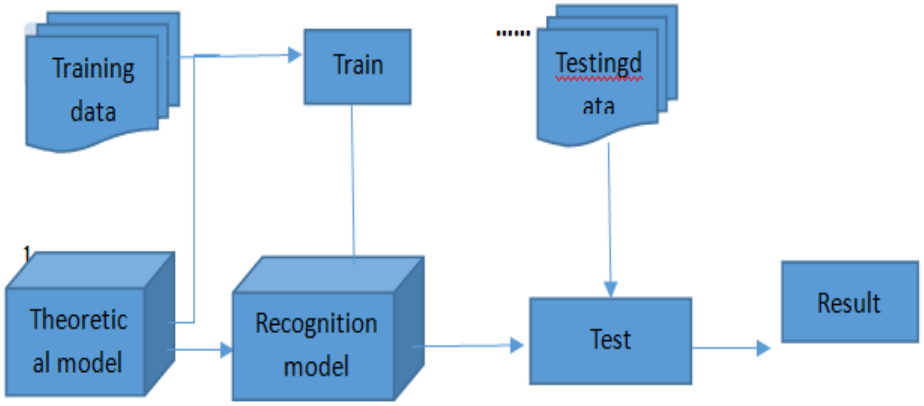


Figure 1. Classification framework based on SVM

Theoretical model

1) Collector: Using port mirroring method to collect data from routers and collect data as raw data and preprocess them. Multiple harvesters can be connected in parallel or in series.

2) Analyzer: The raw data preprocessed by the collector is subjected to a data feature extraction module to extract the characteristic function parameters. Stored in data warehouse. An analyzer can analyze the data of multiple collectors. After the data is preprocessed, the grid search method can be used to verify the optimal parameters of the RBF kernel function for the training data set. So that the analyzer can accurately predict unknown data.

3) After the optimal parameters are determined, the training data set can be trained to obtain the support vector machine model. The extracted parameter data is taken as the feature value of the original data, and the continuous features and discrete features existing in the data are converted, and these heterogeneous data sets are translated into machine-readable values by the data preprocessing module.

4) Multidimensional support vectors are generated by the data after SVM training. At the same time, the multidimensional support vectors are formed through the process of different P2P traffic data, and one support vector library is formed.

5) Known P2P traffic can get specific P2P type through SVM library. Unknown P2P traffic will be subjected to data preprocessing and SVM training by the data acquisition device and analyzer extraction feature extraction module, and the extracted feature information will be added to the SVM support vector library. After obtaining the specific name of the traffic, it is put into the SVM support vector library and finally identifies the specific P2P traffic.

The initial SVM support vector library is a vector library that is trained by known traffic. When the known traffic is subjected to initial data acquisition and feature extraction, data preprocessing, and SVM training, multidimensional support vectors are generated, multidimensional support vectors are characterized, and their characteristic information is added to the SVM support vector library. Known traffic can also form a multidimensional support vector group through the above process.

III. EXPERIMENTS

A. *Traffic data collection*

Select a network server outlet network traffic to carry on the simulation experiment, take 10ms as the sampling time, select the total number of data packets, uplink traffic ratio, average length, TCP traffic ratio and the ratio of the number of connections and different IP number five traffic characteristics as input data feature information, set up the data set as a training sample set and separate and collate, and preprocess the collected data and normalize it. The collected data samples are shown in Table 1.

TABLE I. COLLECTED DATA SAMPLES

DataSet	Time (ms)	Total flow
DataSetA	1hour	2300
DataSetB	1 hour	3020
DataSetC	1 hour	1831
DataSetD	1 hour	2290
DataSetE	1 day	9538

Among them, the first four sets of data are used as input data for the training module. DataSetE is used as the data set to be tested. Three support vector machines are constructed here, namely SVM1, SVM2, and SVM3. After training the classifiers SVM1, SVM2, and SVM3, DataSetE was used as the test sample data set, and experimental results were obtained through the SVM classifier.

B. Finding optimal parameters

The algorithm based on the cross-validation idea is used to select an optimal parameter value C for the RBF kernel function and optimal parameters C and R for the training data set. The labels of the two categories are -1 and 1, which are iterated 51 times. The trained model is saved in the data. Model file. The following information can be obtained from this file: The svm type used for training is c_svm, the kernel function is the radial basis function RBF, the R value is 0.5, the total number of support vectors is 43, and the value of the decision function constant term B is 0.421. Each type of support vector is 22, 20, 21. After the training is completed, the model can be used for SVM type prediction.

Read the file to be predicted, the model file, and then call the function prediction and output the result to a file.

1) After cross test the data, the prediction accuracy is 99.31%.

2) When choose the best parameters (C, R), If the cross validation method of grid search is not adopted, the result of cross validation is not adopted with the default value of 1. According to the method described above, the prediction accuracy is 93.31% obtained by predicting the unknown data through the obtained model. It can be seen that the choice of optimal parameters (C, R) can improve the prediction accuracy of the results.

3) Repeated training and learning. In order to reflect the learning process of SVM, a total of 10 experiments were conducted, by continuously capturing data, the captured data are preprocessed, trained, and predicted. With continuous learning, the accuracy of predictions continues to increase, reaching 91.12%, 93.42%, 94.67%, 95.34%, 95.56%, 96.78%, 97.12%, 97.23%, 97.31% and 97.65% respectively. It can be seen that multiple learning is conducive to classification judgment. However, the learning process also needs to be controlled. Excessive learning will bring negative effects on classification.

IV. DISCUSSION

The model obtained after training can be used for SVM traffic identification. Various P2P traffic and accuracy are identified from packet capture, preprocessing, recognition, learning and training, and compared with the recognition accuracy based on the Bayesian traffic identification model, the recognition method of the SVM has obtained higher accuracy than the original traffic recognition method in practical application. Figure 2 shows the comparison of different traffic models.

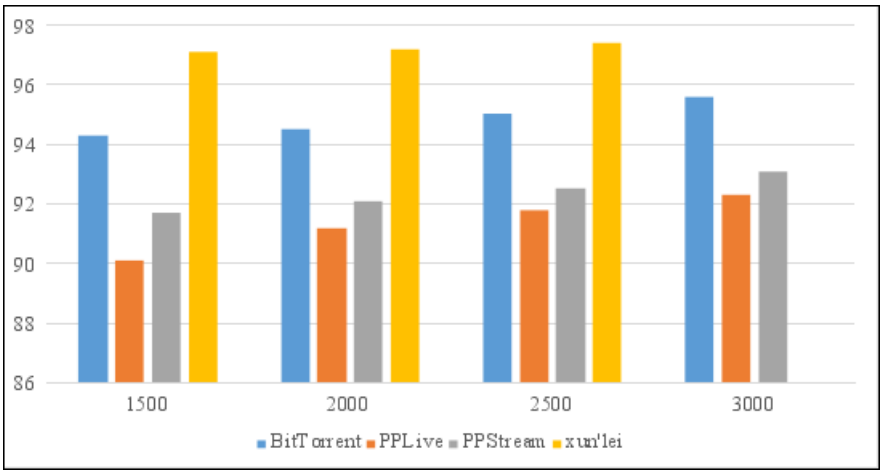


Figure 2. Comparison of different models

From Figure 2, we can see that for the four kinds of P2P traffic in this experiment, the classification and recognition rate of this classifier is all above 90%, so the effect of this MC-SVM classifier on application layer classification of P2P traffic is very good.

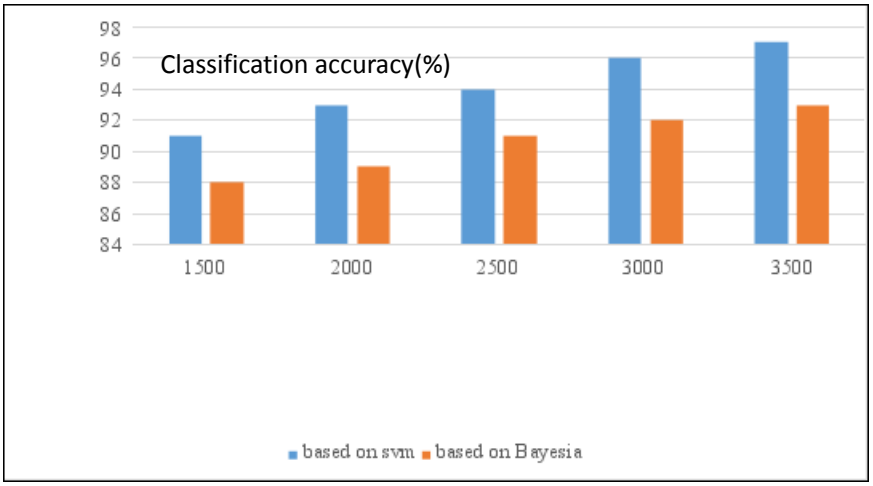


Figure 3. Comparison of stability between Bayesian and SVM

Figure 3 is by using a P2P traffic recognition model based on Bayesian and SVM. With the increase of training data sets, the average classification accuracy can still maintain a certain stability, and the accuracy of recognition reaches 97. 8%. It can be seen that the recognition method of SVM has higher accuracy than the original traffic recognition method in practical application.

V. CONCLUSIONS

SVM algorithm is suitable for nonlinear time series modeling and prediction, so it can well identify the trend of network traffic changes. This paper conducts empirical experiments on the actual data of network traffic. The results show that, compared with the commonly used prediction methods, the recognition model based on SVM can solve the traffic identification. At the same time, it can identify the unknown and large traffic P2P types, and has good effect on the identification of encrypted P2P traffic, and has higher prediction accuracy and better adaptability.

ACKNOWLEDGEMENTS

Fund support: National natural science foundation, Research on Sub-Nyquist Sampling of short pulses based Xampling under Gabor Frames, [61501493]; Shaanxi Education Department Special Fund, Project number: Shaanxi Education Finance [2013] 23.

REFERENCES

[1] Schulze H, Mochalski K. Internet study 2007[J]. Ipoque Gmbh, 2007.

[2] Madhukar A, Williamson C. A Longitudinal Study of P2P Traffic Classification[C]// IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems. IEEE, 2006:179-188.

[3] Ma J, Levchenko K, Kreibich C, et al. Unexpected means of protocol inference[C]// ACM SIGCOMM Conference on Internet Measurement. ACM, 2006:313-326.

[4] Moore A W, Papagiannaki K. Toward the accurate identification of network applications[C]// International Conference on Passive and Active Network Measurement. Springer-Verlag, 2005:41-54.

[5] Haffner P, Sen S, Spatscheck O, et al. ACAS: automated construction of application signatures[C]// ACM SIGCOMM Workshop on Mining Network Data. ACM, 2005:197-202.

[6] Huang K, Zhang Q, Zhou C, et al. An Efficient Intrusion Detection Approach for Visual Sensor Networks Based on Traffic Pattern Learning[J]. IEEE Transactions on Systems Man & Cybernetics Systems, 2017, PP(99):1-10.

[7] Yuan R, Li Z, Guan X, et al. An SVM-based machine learning method for accurate internet traffic classification[J]. Information Systems Frontiers, 2010, 12(2):149-156.

[8] I ANA. Internet assigned numbers authority[EB／OL]. http：//www.iana.org/assigu mens/port-numbers.

[9] Spatscheck O, Sen S, Wang D. Method and apparatus for automatically constructing application signatures: US, US 7620807 B1[P]. 2009.

[10] Yang C S, Liao M Y, Luo M Y, et al. A Network Management System Based on DPI[C]// International Conference on Network-Based Information Systems. IEEE Computer Society, 2010:385-388.

[11] Hartigan J A, Wong M A. Algorithm AS 136: A K-Means Clustering Algorithm[J]. Journal of the Royal Statistical Society, 1979, 28(1):100-108.

[12] Liu H, Yu L. Yu, L.: Toward Integrating Feature Selection Algorithm for Classification and Clustering. IEEE Transaction on Knowledge and Data Engineering 17(4), 491-502[J]. IEEE Transactions on Knowledge & Data Engineering, 2005, 17(4):491-502.

[13] Pedro S D S. Collective intelligence as a source for machine learning self-supervision[C]// International Workshop on Web Intelligence & Communities. ACM, 2012:5.

[14] Chapelle O. Semi-supervised Learning (Adaptive Computation and Machine Learning)[J]. Mit Pr, 2006.

[15] Liu S, Dou Z T, Li F, et al. A new ant colony clustering algorithm based on DBSCAN[C]// International Conference on Machine Learning and Cybernetics. IEEE, 2004:1491-1496 vol.3.

[16] Este A, Gringoli F, Salgarelli L. On-line SVM traffic classification[C]// Wireless Communications and Mobile Computing Conference. IEEE, 2011:1778-1783.

[17] Osuna E, Freund R, Girosi F. Training svm: An application to face detection[C]// 1997.

[18] Este A, Gringoli F, Salgarelli L. On-line SVM traffic classification[C]// Wireless Communications and Mobile Computing Conference. IEEE, 2011:1778-1783.

[19] Zhou X. A P2P Traffic Classification Method Based on SVM[C]// International Symposium on Computer Science and Computational Technology. IEEE Computer Society, 2008:53-57.

[20] Bernaille L, Teixeira R, Akodkenou I, et al. Traffic classification on the fly[J]. Acm Sigcomm Computer Communication Review, 2006, 36(2):23-26.

[21] JinHuaXu, HongLiu. Web User Clustering Analysis based on KMeans Algorithm[C]// 2010 international conference on information,networking and automation. 2010:V2-6-V2-9.

[22] Poornalatha G, Raghavendra P S. Web User Session Clustering Using Modified K-Means Algorithm[M]// Advances in Computing and Communications. Springer Berlin Heidelberg, 2011:243-252.

[23] Wang T Z. The Development of Web Log Mining Based on Improve-K-Means Clustering Analysis[M]// Advances in Computer Science and Information Engineering. Springer Berlin Heidelberg, 2012:613-618.

[24] Ye M, Wu J, Xu K, et al. Identify P2P Traffic by Inspecting Data Transfer Behaviour[J]. Computer Communications, 2010, 33(10):1141-1150.

[25] Tapaswi S, Gupta A S. Flow-Based P2P Network Traffic Classification Using Machine Learning[C]// International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery. IEEE, 2013:402-406.

[26] Deng H, Yang A M. P2P traffic classification method based on SVM[J]. Computer Engineering & Applications, 2006.

[27] Yang A M, Jiang S Y, Deng H. A P2P Network Traffic Classification Method Using SVM[C]// Young Computer Scientists, 2008. Icycs 2008. the, International Conference for. IEEE, 2008:398-403.

[28] Jiang W, Wang C Z, Luo H F, et al. Research on a Method of P2P Traffic Detection Based on SVM[J]. Journal of Hubei University of Technology, 2010.

[29] Zhu A. A P2P Network Traffic Classification Method Based on C4.5 Decision Tree Algorithm[M]// Proceedings of the 9th International Symposium on Linear Drives for Industry Applications, Volume 4. Springer Berlin Heidelberg, 2014:373-379.

Multi Antenna Precoding Algorithm Based on M Spread Spectrum

Sun Ruihua
College of Information Engineering
North China University of Science and Technology
Tangshan, Hebei, China
e-mail: 18712850210@163.com

Bai Junying
College of Information Engineering
North China University of Science and Technology
Tangshan, Hebei, China
e-mail: 15100586578@163.com

An Yongli*
College of Information Engineering, North China University of Science and Technology
Tangshan, Hebei, 063009, China
***Corresponding Author** tongxinayl@126.com

***Abstract*—MIMO multi antenna technology can increase the capacity and channel utilization of the communication system without increasing the bandwidth, and become the key technology in the new generation of mobile communication system. However, each channel has its own channel parameters, so in the process of signal transmission, the influence of channel parameters should be considered. When the signal is received, it needs to be restored, which leads to the complexity of the receiving signal. Therefore, this paper proposes a multi antenna precoding algorithm based on M spread spectrum, precoding before sending signal spread spectrum to simplify the signal receiving equipment, and verify the feasibility of the algorithm through system error rate.**

Keywords-MIMO; Multi Antenna Technology; Mobile Communication; Spread Spectrum; Precoding

I. INTRODUCTION

With the rapid popularization of the practical application of wireless communication system, the number of wireless communication users and user service demand have increased exponentially, but the radio spectrum resources which can be used in wireless communication services are extremely limited. The contradiction between the increasing demand of wireless service and the limited radio spectrum resources is becoming more and more prominent. MIMO multi antenna technology is a new type of wireless communication technology developed under this background. Multi antenna [1] can effectively improve the channel capacity and is widely used. Multi antenna technology can make full use of space resources to achieve multiple and multi harvest. Without increasing the spectrum resources and transmitting power of the antenna, it can increase the capacity of the system channel, and has obvious advantages in many technologies. It is regarded as the core technology of the next generation mobile communication.

However, the influence of channel parameters on receiver terminals can not be ignored in signal transmission. Therefore, a precoding algorithm based on M spread spectrum is proposed, which is pre processed before sending signals. In order to achieve the purpose of simplifying the receiving device.

II. MULTI ANTENNA MIMO SYSTEM

A. Multi Antenna MIMO System Model

MIMO multi antenna technology is a major breakthrough in antenna technology in the field of wireless mobile communication. In theory, it can improve the system capacity and frequency efficiency without increasing the time and frequency. Its concept is very simple. It needs a transmitter and a receiver that have multiple antennas to carry out signal transmission simultaneously, so that a wireless MIMO system can be formed. Figure 1 is a schematic block diagram of the MIMO system:

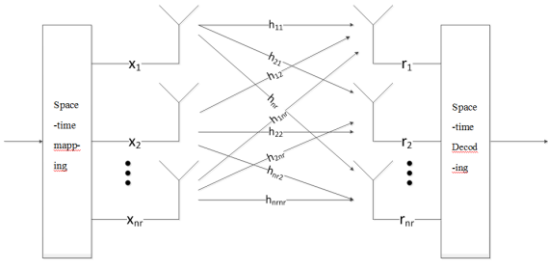


Figure 1. MIMO system schematic diagram

The transmitter mapped the data signal sent by the space-time map to multiple antennas and sent them out. The receiver sent the signals received by all the antennas to the space-time decoding to restore the data signals sent by the transmitting terminal. According to the difference of space-time mapping, MIMO technology can be roughly divided into two categories: spatial diversity and spatial multiplexing. Spatial diversity refers to the use of multiple transmission antennas to send signals with the same information through different paths, and at the same time, multiple independent fading signals of the same data symbol are obtained at the receiver end, thus obtaining the reliability of the diversity

enhancement. For example, in the slow Rayleigh fading channel, using a transmitting antenna, the N root receiving antenna sends signals through N different paths. If the fading between antennas is independent, the maximum diversity gain can be N. For transmitter diversity technology, the gain of multiple paths is also used to improve the reliability of the system. In a system with the N root receiving antenna with the M root transmitting antenna, the maximum diversity gain of M*N can be obtained if the path gain between the antenna pairs is an independent uniform distribution of Rayleigh fading. At present, the commonly used spatial diversity techniques in MIMO systems are Space Time Block Code (STBC) and beamforming technology. STBC is an important coding form based on transmit diversity, the most basic of which is the Alamouti design for two antennas.

The signal model of MIMO multi antenna system:

$$\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_{Nr} \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1Nt} \\ h_{21} & h_{22} & \cdots & h_{2Nt} \\ \vdots & \vdots & \vdots & \vdots \\ h_{Nr1} & h_{Nr2} & \cdots & h_{NrNt} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{Nt} \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_{Nt} \end{pmatrix} \quad (1)$$

The matrix is $r = Hx + n$, r is the received signal, H is the channel matrix, x is the transmit signal, and n is the noise signal.

The transmitter mapped the data signal sent by the space-time map to multiple antennas and sent them out. The receiver sent the signals received by all the antennas to the space-time decoding to restore the data signals sent by the transmitting terminal. According to the difference of space-time mapping, MIMO technology can be roughly divided into two categories: spatial diversity and spatial multiplexing. Spatial diversity refers to the use of multiple transmission antennas to send signals with the same information through different paths, and at the same time, multiple independent fading signals of the same data symbol are obtained at the receiver end, thus obtaining the reliability of the diversity enhancement. Spatial multiplexing technology uses multiple antennas to transmit independent data at the same time, thus increasing the data capacity of the system.

For example, in the slow Rayleigh fading channel, using a transmitting antenna, N root receiving antennas sends signals through N different paths. If the fading between antennas is independent, the maximum diversity gain can be N. For transmitter diversity technology, the gain of multiple paths is also used to improve the reliability of the system. In a system with the N root receiving antennas and with the M root transmitting antennas, the maximum diversity gain of M*N can be obtained if the paths gain between the antenna pairs is an independent uniform distribution of Rayleigh fading.

The diversity technique is mainly used to combat channel fading. Conversely, the fading characteristics of MIMO channels can provide additional channels to increase the degree of freedom in communication. In essence, if each fading between the transmit and receive antennas is independent, multiple parallel subchannels can be generated.

If we transmit different information streams on these parallel sub channels, we can provide transmission data rate, which is called spatial multiplexing. According to the correspondence between sub data stream and antenna, the spatial multiplexing system can be roughly divided into three modes: D-BLAST, V-BLAST and T-BLAST.

B. Main Teachnolody

There are three main technologies in MIMO system: space multiplexing, transmission diversity and beamforming.

1) Space reuse:

The system divides the data into multiple parts, and the system is transmitted on the multiple antennas at the transmitter. After receiving the mixed signals of multiple data, the parallel data streams are distinguished by the independent fading characteristics between different space channels. It achieves the purpose of obtaining higher data rate in the same frequency resource.

2) Transmission diversity technology:

Taking the space time coding as the representative, the data stream is jointly encoded at the transmitter side to reduce the symbol error rate due to channel fading and noise. Space time coding increases the redundancy of the signal at the transmitter, so that the diversity gain is obtained at the receiver.

At present, the commonly used spatial diversity techniques in MIMO systems are Space Time Block Code (STBC) and beamforming technology. STBC is an important coding form based on transmit diversity, the most basic of which is the Alamouti design for two antennas.

3) Beamforming:

The system generates a directivity beam through multiple antennas, concentrating the signal energy in the direction of the desired transmission, thus improving the quality of the signal and reducing the interference to other users.

Space reuse can maximize the average transmission rate of MIMO system, but only a limited diversity gain can be obtained. It may not be used in high order modulation, such as 16QAM, in the use of SNR.

Wireless signals will be reflected frequently in dense urban areas, indoor coverage and other environments, making the fading characteristics of multiple spatial channels more independent, thus making the effect of space multiplexing more obvious. Wireless signals are less in the suburbs and in rural areas, and the correlation between different spatial channels is larger, so space reuse is therefore reused. Which effect is much worse. The extra diversity gain and coding gain can be obtained by space-time coding of the transmitted signal, so the high order modulation can be used in the wireless environment with relatively small SNR, but the rate bonus of the space parallel channel can not be obtained. Space coding technology also performs well in situations where wireless correlation is large.

Beamforming technology can achieve better signal gain and interference suppression when it can acquire channel state information, so it is more suitable for TDD system.

Beamforming is not suitable for dense urban areas, indoor coverage and other environments. Due to reflection, on the one hand, the receiver receives signals from too many

paths, which results in a poor phase effect. On the other hand, a large number of multipath signals will lead to the difficulty of DOA information estimation.

C. The Advantages of MIMO System Model

The application of MIMO technology makes space a kind of resource that can be used to improve performance, and can increase the coverage of wireless system.

1) Improving the capacity of the channel

The MIMO access point can transmit and receive multiple spatial flows between the MIMO access point and the client side. The channel capacity can increase linearly with the increase of the number of antennas. Therefore, the capacity of the wireless channel can be doubled by using the MIMO channel. Without increasing the bandwidth and the transmit power of the antenna, the spectrum utilization rate can be doubled.

2) Improving the reliability of the channel

By using the spatial multiplexing gain and spatial diversity gain provided by MIMO channel, multiple antennas can be used to suppress channel fading. The application of multi antenna system enables parallel data stream to be transmitted at the same time, which can significantly overcome the channel fading and reduce the bit error rate.

D. Application

1) Wireless broadband mobile communication

The wireless broadband mobile communication system with MIMO technology can be divided into two categories from the multi antenna placement method of the base station. One is that multiple base station antennas are arranged to form an antenna array and are placed in the coverage area. This class can be called a centralized MIMO, and the other is that the multiple antennas of the base station are scattered in the coverage area. It is called a distributed MIMO.

2）Traditional cellular mobile communication system

MIMO technology can be applied directly to traditional cellular mobile communication systems, and the single antenna of base stations can be changed into antenna arrays. The base station carries out MIMO communication with the mobile station with multiple antennas in the cell through the antenna array.

3）Combining with the traditional distributed antenna system

The combination of traditional distributed antenna system and MIMO technology can improve the capacity of the system. This new distributed MIMO system structure, distributed wireless communication system (DWCS), has become an important research focus of MIMO technology.

4） The field of wireless communication

MIMO technology has become one of the key technologies in the field of wireless communication. Through the continuous development in recent years, MIMO technology will be more and more applied to all kinds of wireless communication systems.

5）Radar field

MIMO technology is also used in the field of radar. It mainly uses multiple antennas to transmit different orthogonal waveforms, and covers large space at the same time, and uses long time coherent accumulation to obtain high signal to noise ratio.

III. SPREAD SPECTRUM COMMUNICATION

A. Spread Spectrum Communication technology

The spread spectrum communication technology [5] is a way of information transmission. The bandwidth of the signal is far greater than the minimum bandwidth required for the information transmitted; the expansion of the frequency band is accomplished by an independent code sequence, implemented by the encoding and modulation methods, and is independent of the information data; the same code is used at the receiver. Related synchronous reception, expansion and recovery of transmitted information. The spread spectrum code is used to spread spectrum modulation at the sending end, and the correlation demodulation technology is used to receive the signal at the receiver.

Spread spectrum communication needs spread spectrum modulation to transmit spread spectrum modulation, and signal reception needs to be extended with the same spread spectrum coding, which provides the basis for frequency multiplexing and multiple access communication. Making full use of the correlation characteristics between the spread spectrum codes of different code types, it can be allocated to different users and different spread spectrum codes, which can distinguish different users' signals and do not be disturbed by other users, and the frequency reuse can be realized.

Spread spectrum signal is obtained by spreading the random sequence pseudo-random code to modulate radio frequency signal or to jump the frequency of carrier signal. Therefore, the spread spectrum system is different from the traditional communication system, and it can share the same channel resources to the maximum extent. Each system has a different extension sequence to reduce interference from other devices. Only recipients with the same extension sequence with the transmitter can restructure or compress the spread spectrum signal to obtain effective loading information. Even if a set of spread spectrum devices use the same channel to transmit signals in the same area, they will not interfere with each other if they use different spread spectrum sequences. The advantage of the channel reuse of spread spectrum system makes it the most ideal choice in the crowded environment of big cities.

B. Spread Spectrum Principle

At the transmitter, the input information is first modulated by the information to form a digital signal, and then the spread spectrum code sequence generated by the spread spectrum code generator is used to modulate the digital signal to broaden the spectrum of the signal. The broadened signal is then modulated to radio frequency. At the receiving end, the received wideband radio frequency signal is converted to the intermediate frequency, and then the local spread spectrum code sequence generated from the same origin is despreading, and then the information is

demodulated into the original information output. Figure 2 is a schematic map of spread spectrum technology.

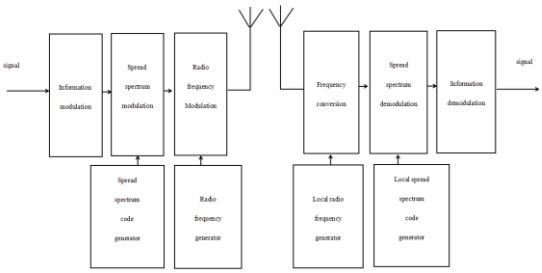


Figure 2. Principle diagram of spread spectrum

1) Transmitting terminal

i) The information input from the transmitter is modulated by information to form a digital signal.

ii) Spread spectrum code generated by spread spectrum code generator to expand the spectrum of digital signal.

iii) The digital signal of RF generator is converted into analog signal and sent through RF signal.

2) Receiving *termina*l

i) At the receiving end, the received RF signals are converted from high frequency to intermediate frequency that can be processed by electronic devices, and the analog signals are converted into digital signals.

ii) The spread spectrum code generator produces the same spread spectrum code as the sending end to despread the digital signal.

iii) Demodulating the digital signal into the original information output.

C. Classification of Spread Spectrum Technology

In technical implementation, spread spectrum is usually divided into several methods: direct sequence (DS) spread spectrum, frequency hopping (FH) spread spectrum, time hopping (TH) spread spectrum and linear frequency modulation (Chirp) spread spectrum.

1) Direct sequence spread spectrum

The spread spectrum sequence with high bit rate is used to expand the spectrum of the signal at the transmitter. At the receiver, the same spread spectrum sequence is used to despread, and the spread spread spectrum signal is restored to original information.

2) Frequency hopping spread spectrum

Multiple frequency shift keying is selected by using a sequence of codes. That is to say, the frequency shift keying modulation using the spread spectrum code sequence makes the carrier frequency jump.

3) Time hopping spread spectrum

Cause the signal to jump on the time axis. First, the time axis is divided into many time pieces, which is controlled by the sequence of spread spectrum code in one frame. In other words, the time jump can be understood as the time shift keying of the multi time slice selected by a certain code sequence.

4) Linear frequency hopping

The transmitted radio frequency pulse signal is broadened in one cycle, and the spread spectrum modulation method is mainly applied to radar.

D. Application of Spread Spectrum Communication

As a mature high-tech technology, spread spectrum communication can be applied to:

(1) The dilute rural areas and underdeveloped areas of the remote people;

(2) The prosperous downtown area of the Saturated wired infrastructure;

(3) New communities with cable infrastructure lagging due to surging business requirements;

(4) User backbone / backup communication network to make up for the shortage of public network of Posts and telecommunications.

IV. PRECODING ALGORITHM BASED ON M SPREAD SPECTRUM

A. Pseudorandom Code Theory

Pseudo random code (Pseudo Random Code, Pseudo Noise Code, PN code, pseudo-noise code) is a code with a similar white noise character, also known as a random (pseudo-noise) sequence. The structure can be pre determined, and can be repeatedly generated and copied, with a random sequence of random characteristics. Pseudorandom code sequences can be generated by the shift register network. The network consists of a RP cascade dual state device shift pulse generator and a modular two adder. White noise is a random process, the instantaneous value obeys the normal distribution, and the power spectrum is uniform in a wide band. With excellent correlation characteristics, the autocorrelation function of white noise is similar to the delta function. But it can not realize amplification, modulation, detection, synchronization and control.

Most pseudo random codes are periodic codes, which can be generated and copied artificially, usually generated by binary shift registers. With the nature of white noise, the correlation function has a sharp characteristic, the power spectrum occupies a very wide band, so it is easy to separate from other signals or interference with excellent anti-interference characteristics.

In engineering, pseudo-random codes are commonly used to represent pseudo random codes in two yuan domain 0, 1, 0 and 1 elements.

(1) In each cycle, the number of 0 elements and 1 elements is approximately equal, and the maximum is only one difference.

(2) Within each cycle, the number of element runs of k bit length appears more than twice as many times as the length of k+1 bits (the same element of the same r bit that appears continuously) is called the element distance of the length of the r bit).

(3) The autocorrelation function of a sequence is a periodic function and has a dual value property.

$$R(\tau)=\begin{cases}1 & \tau=mN\\ -\dfrac{k}{N} & \tau\neq mN\end{cases}\qquad m=0,\pm1,\pm2,.... \tag{2}$$

In the formula, N is the cycle of two yuan sequence, also known as code length or length; k is integer less than N; τ is symbol delay.

Pseudo-random codes have the following characteristics:

(1) The pseudo random signal must have sharp autocorrelation function, and the cross-correlation function value should be close to 0 value.

(2) There is enough code cycle to ensure the requirements of anti detection and anti-jamming.

(3) The number of codes is enough to be used as independent addresses to achieve code division multiple access requirements.

(4) It is easy to be produced in engineering. Birth, processing, reproduction and control.

Setting { ai } and { bi } is the two code sequence of N,so $a_{N+i}=a_i \quad ,b_{N+i}=b_i$.

Cross correlation function:

$$R_{ab}(\tau)=\frac{1}{N}\sum_{i=1}^{N}a_i b_{i+\tau} \tag{3}$$

If $R_{ab}(\tau)=0$, then ai is orthogonal to bi.

Autocorrelation function:

$$R_a(\tau)=\frac{1}{N}\sum_{i=1}^{N}a_i a_{i+\tau} \tag{4}$$

1) Narrow sense pseudorandom sequence

If the length of the code is N, the autocorrelation function of the { ai } sequence is

$$R_a(\tau)=\frac{1}{N}\sum_{i=1}^{N}a_i a_{i+\tau}=\begin{cases}1 & \tau=mN\\ -\dfrac{1}{N} & \tau\neq mN\end{cases}\qquad m=0,\pm1,\pm2,... \tag{5}$$

2) Generalized pseudorandom sequence

If the length of the code is N, the autocorrelation function of the { ai } sequence is

$$R_a(\tau)=\frac{1}{N}\sum_{i=1}^{N}a_i a_{i+\tau}=\begin{cases}1 & \tau=mN\\ \alpha<1 & \tau\neq mN\end{cases}\qquad m=0,\pm1,\pm2,... \tag{6}$$

B. M Sequence

The m sequence is a pseudo random sequence, pseudo noise (PN) code or pseudo-random code. A sequence that can be determined and can be repeated is called deterministic sequence. A sequence of random sequences that cannot be determined in advance and can not be repeated. Sequences that cannot be predefined but can be repeated are called pseudo random sequences. The m sequence is a code sequence whose the cycle is $2\wedge n-1$ generated by a n-linear shift register, which is the abbreviation of the longest linear shift register sequence.

For a n level feedback shift register, there can be up to 2^n states. For a linear feedback shift register, the full "0" state will not be transferred to other states, so the longest period of the sequence of the linear shift register is $2\wedge n-1$.When the period of the {ai} sequence generated by the n level linear shift register is $2\wedge n-1$, {ai} is called a n class m sequence. When the feedback function is a nonlinear function, a nonlinear shift register is formed, and its output sequence is nonlinear sequence. The maximum cycle of output sequence can reach 2^n, and the nonlinear shift register sequence with the maximum cycle value is called m sequence.

Generally speaking, in a n level binary shift register generator, the maximum length of code generation cycle is $2\wedge n-1$. Take m=4 as an example, if its initial state is (a3,a2,a1,a0)=(1,0,0,0), then a new input a4 =1$\oplus$ 0=1 is generated by a3 and a0 mode 2 at the time of shift, and the new state becomes (a3,a2,a1,a0) = (1,0,0,0) , so that the shift returns to the initial state 15 times, but if the initial state (0, 0, 0, 0), Then, after the shift, the whole state is 0, which means that the whole 0 state should be avoided in this feedback. There are 24=16 different states in the 4 stage. There are 15 kinds of availability except all 0 states, that is, the maximum period of the sequence generated by any 4 level feedback latch is up to 15, which satisfies the 2n-1.

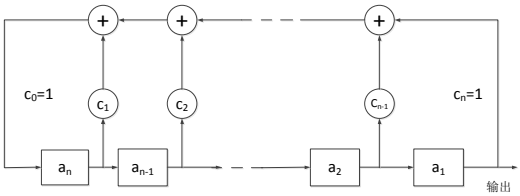


Figure 3. N class linear feedback latch

ai(i=0~n) - the state of the latch. ai=0 or 1 - feedback state. ci=0 indicates that the feedback line is disconnected, and ci=1 means the feedback line is connected.

Figure 3 shows the composition of a general pure feedback latch. The connection state of the feedback line is expressed in ci, ci=1 indicates that the line is connected, the ci=0 is disconnected, and the connection state of the feedback line is different, which may change the period of the latch.

In order to generate m sequences, the characteristic polynomial must be determined to determine the feedback structure of the linear feedback shift register. The characteristic equation of the N class linear shift register is defined as:

$$f(x)=1\oplus c_1x\oplus c_2x^2\oplus\cdots\cdots\oplus c_nx^n \tag{7}$$

The original polynomial of the m sequence is as follows: $A(x)=x^5+x^2+1$, Figure 4 is a shift register structure diagram.

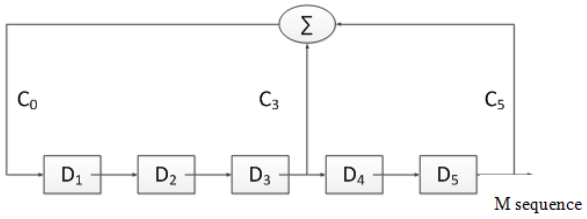


Figure 4. Shift register structure diagram

The initialization register is [D5 D4 D3 D2 D1]=[0 0 0 0 1], the register first shifts left to see m (0) =0, and then according to the above picture, we can see feedback D1=C5 $\oplus$ C3. Because of the 5 order register, the code length is N=25-1=31. So 31 cycles are needed to get the required m sequence. The simulation results are as follows:

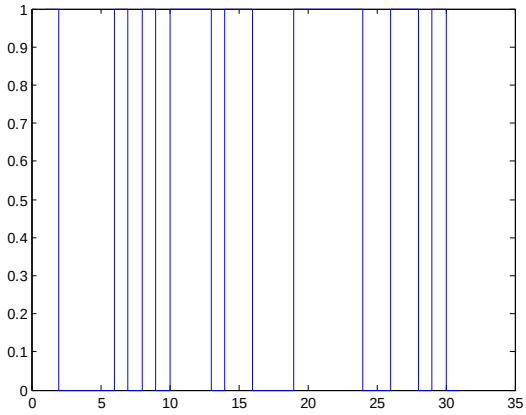


Figure 5. M sequence simulation result diagram

C. The Properties of M Sequence

1) Equilibrium

In a period of m sequence, the number of symbols "1" and "0" are roughly equal, "0" appears 2n-1-1 times, and "1" appears 2n-1 times ("1" more than "0").

2) Run length distribution

Run length refers to the same element in the sequence. And the number of this element is called the length of the travel.

3) Shift additive properties

A m sequence Mp is added to a different sequence of Mr ,generated by any delay shift, modules 2 and is still a Mz of a delay shift sequence of Mp, that is, Mp $\oplus$ Mr=Mz. Now, the m sequence of a m=7 is now taken as an example, One period of Mp is set to 1110010, and the other sequence Mr is the result that Mp moves to the right one time, that is, a corresponding period of Mr is 0111001, the two sequence modules 2 and the corresponding period of the 1110010 $\oplus$ 0111001=1001011 upper form for Mz, which is the same as the result of the Mp shift to the right 5 times.

4) Autocorrelation characteristics

Autocorrelation and cross correlation:

A m sequence and its shifted sequence are 2 bit by bit, the sequence obtained is also a m sequence, but the phase is different. The m sequence for 2 different phases in the m sequence generator. When the period p is large and the r module is $p\neq 0$, the two sequences are almost orthogonal.

$$R(j)=\begin{cases}1, & j=0\\ \dfrac{-1}{m}, & j=1,2,\cdots,m-1\end{cases} \tag{8}$$

5) Periodicity

As the m sequence has periodicity, its autocorrelation function is also cyclical, and the period is m, namely $R(j)=R(j-km)$, when $j\geq km,k=1,2,\ldots$

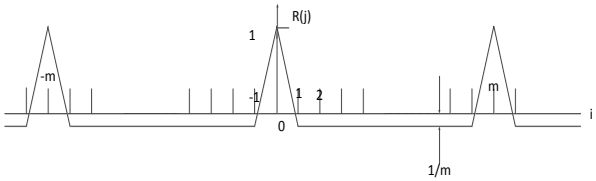


Figure 6. Periodic schematic diagram of m sequence

The maximum length of the m sequence depends on the progression of the shift register, and the structure of the code depends on the location and quantity of the feedback. Different tapped combinations produce code sequences of different lengths and structures, and some tap combinations fail to produce the longest cycle sequences. A great deal of research has been done on what kind of length and sequence of code can be produced by tap. The connection diagram of the 100 level M sequence generator and the structure of the generated m sequence have been obtained.

6) Power spectral density

Power spectral density and autocorrelation coefficient constitute a pair of Fu Liye transform. Find out as follows:

$$R(\omega)=\frac{m+1}{m^2}\left[\frac{\sin(\omega T/2m)}{\omega T/2m}\right]^2\sum_{n\to\infty}^{\omega}\delta\left(\omega-\frac{2\pi n}{T}\right)+\frac{1}{m^2}\delta(\omega) \tag{9}$$

Because when m is large, the equilibrium of the m sequence, the range distribution, the autocorrelation and the power spectrum density are all similar to the white noise, but it has the regularity and can be repeated, so the m sequence belongs to a pseudo noise sequence.

D. Application of M Sequence and Its Significance

1) Application in communication encryption

The autocorrelation of m sequence is good, it is easy to produce and copy, and has pseudo randomness. Using m sequence to encrypt the digital signal, the encrypted signal has the characteristic of pseudo noise while carrying the original signal, so as to achieve the purpose of hiding information in the process of signal transmission; at the receiver, the m sequence is used again to decrypt and restore the original signal.

2) The application of the radar signal design

In recent years, the signal used in the spread spectrum radar is a pseudo random sequence with a modulated noise character. It has a high distance resolution and velocity resolution. The receiver of this radar works by means of correlation demodulation. It can work at low SNR and has strong anti-interference capability. The radar is a kind of continuous wave radar, which has low probability of interception. It is a kind of new radar, high performance and suitable for modern high-tech war. The radar system using pseudo-random sequences as launching signals has many outstanding advantages. First, it is a continuous wave radar, which can make good use of transmitter power. Secondly, in a certain signal to noise ratio, it can achieve a good measurement accuracy and guarantee the single value of the measurement, which has a higher distance resolution and velocity resolution than the monopulse radar. Finally, it has strong anti-jamming ability, and the enemy will interfere with this wideband radar signal, which will be much more difficult than interfering with ordinary radar signals.

3) Application in communication system

Pseudo random sequence is a seemingly random, actually regular periodic binary sequence, which has the properties similar to noise sequence. In CDMA, the address code is selected from pseudo random sequence, and a pseudo random sequence is used most easily in CDMA; m sequence is used to distinguish different users from different phases of the m sequence. For data security, a data mask (data disruption) technique is used in the paging channel and forward service channel of the CDMA, which is used to scramble the traffic channel with the m sequence of the length of $2^{42}-1$, which is performed on the modulation characters output by the packet interleaver. Through the interleaver output character and the long code PN chip, the binary mode addition is completed.

E. Precoding Algorithm Based on M Spread Spectrum

On the basis of the original spread spectrum technology, the algorithm proposed a new technology. The general signal must have the matrix parameters of the channel itself during a certain channel during the transmission and reception. In this way, the reduction of the signal is difficult when receiving the signal at the receiving end. In other words, in real life, the general signal sending and receiving may make the device of the receiver complex. In order to simplify the receiving device, the inverse matrix of the channel is multiplied before the signal is sent, so that the purpose is achieved within a certain range of bit error rate.

The design steps are as follows:

Step (1): The data flow of the first user is $\{b_1^{(k)}\}$, $\{b_2^{(k)}\}$ and $\{b_3^{(k)}\}$. The base station transmitter encodes the data of three channels to get coded signals $s_1^{(k)}$, $s_2^{(k)}$ and $s_3^{(k)}$:

$$\begin{aligned} s_1^{(k)} &= b_1^{(k)} G_1 \\ s_2^{(k)} &= b_2^{(k)} G_2 \\ s_3^{(k)} &= b_3^{(k)} G_3 \end{aligned} \tag{10}$$

Among them, G_1, G_2 and G_3 are generating matrices. The three spatial channel of user K uses three different encoding. G_1, G_2, and G_3 corresponding check matrices are H_1, H_2 and H_3;

Step 2): The coding signals $s_1^{(k)}$, $s_2^{(k)}$ and $s_3^{(k)}$ of the K user are modulated by the channel matrix.

$$\begin{bmatrix} \mathbf{z}_1^{(k)} \\ \mathbf{z}_2^{(k)} \\ \mathbf{z}_3^{(k)} \end{bmatrix} = \begin{bmatrix} h_{11}^{(k)} & h_{21}^{(k)} & h_{31}^{(k)} \\ h_{12}^{(k)} & h_{22}^{(k)} & h_{32}^{(k)} \\ h_{13}^{(k)} & h_{23}^{(k)} & h_{33}^{(k)} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{s}_1^{(k)} \\ \mathbf{s}_2^{(k)} \\ \mathbf{s}_3^{(k)} \end{bmatrix} \tag{11}$$

Among them, $h_{ij}^{(k)}, i, j \in \{1,2,3\}$ is the attenuation coefficient of the base station transmitter antenna i to the k mobile receiver antenna j through the independent Rayleigh path. and the baseband modulation signals of the k users, $z_1^{(k)}$, $z_2^{(k)}$ and $z_3^{(k)}$, $k = 1, \ldots, k$ are obtained.

Step 3): The base band transmitter modulates the baseband modulation signals $z_1^{(k)}$, $z_2^{(k)}$ and $z_3^{(k)}$, $k = 1, \ldots, k$, and obtains the signal $t(z_1^{(k)})$, $t(z_2^{(k)})$ and $t(z_3^{(k)})$ based on the M spread spectrum precoding, and then the three antennas are transmitted respectively.

Step 4): The k mobile receiver uses the local despreading circuit to extract the baseband encoded signals $y_1^{(k)}$, $y_2^{(k)}$ and $y_3^{(k)}$.

$$\begin{bmatrix} t^{-1}(r_1^{(k)}) \\ t^{-1}(r_2^{(k)}) \\ t^{-1}(r_3^{(k)}) \end{bmatrix} = \begin{bmatrix} y_1^{(k)} \\ y_2^{(k)} \\ y_3^{(k)} \end{bmatrix} = \begin{bmatrix} h_{11}^{(k)} & h_{21}^{(k)} & h_{31}^{(k)} \\ h_{12}^{(k)} & h_{22}^{(k)} & h_{32}^{(k)} \\ h_{13}^{(k)} & h_{23}^{(k)} & h_{33}^{(k)} \end{bmatrix}^T \begin{bmatrix} z_1^{(k)} \\ z_2^{(k)} \\ z_3^{(k)} \end{bmatrix} + \begin{bmatrix} n_1^{(k)} \\ n_2^{(k)} \\ n_3^{(k)} \end{bmatrix} \quad (12)$$

In formula 12, $n_1^{(k)}$ is the baseband noise vector for the first antenna's channel of the k mobile station receiver, $n_2^{(k)}$ is the baseband noise vector for the second antenna's channel of the k mobile station receiver, $n_3^{(k)}$ is the baseband noise vector for the third antenna's channel of the k mobile station receiver , $t^{-1}(r_1^{(k)})$, $t^{-1}(r_2^{(k)})$ and $t^{-1}(r_3^{(k)})$ are the representations of despreading.

Step 5): The k receiver uses a local decoder to decode the received baseband signals $y_1^{(k)}$, $y_2^{(k)}$ and $y_3^{(k)}$, and extracts the data streams of the base station to transmit data streams $b_1^{(k)}$, $b_2^{(k)}$ and $b_3^{(k)}$, with $b_1^{(k)}H_1^{(k)}=0$, $b_2^{(k)}H_2^{(k)}=0$ and $b_3^{(k)}H_3^{(k)}=0$ without error.

V. SIMULATION ANALYSIS

Taking the 3*3 antenna system as an example, a binary original signal with a sequence length of 9 is sent as shown in Figure 7. Before sending the signal, a binary signal with a spread spectrum growth of 15 is spread out with a precoding algorithm based on M spread spectrum, and the binary signal processed by the algorithm is shown as shown in Figure 8. after the receiver despreading. It is found that the original signal and the received signal have some error. Therefore, the different antenna systems using the algorithm, the error rate simulation under the same signal to noise ratio, and under the same antenna system, the algorithm is applied to simulate the error rate of the algorithm without using the algorithm.

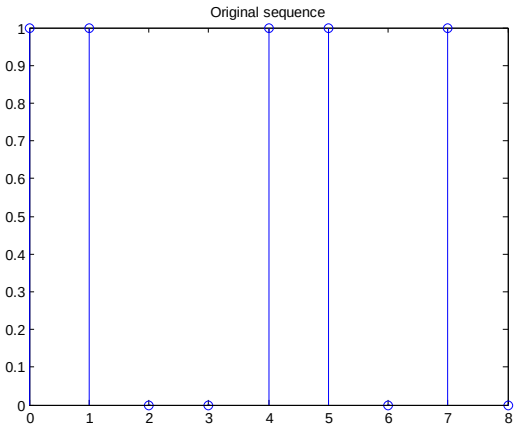


Figure 7. Original signal

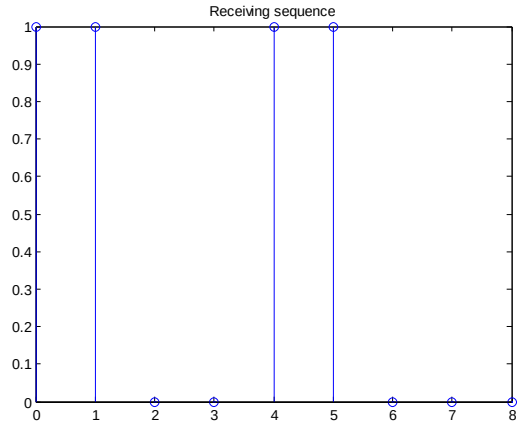


Figure 8. The signal received by this algorithm

The error rate simulation is carried out with different antenna systems. The results are shown in Figure 9. The bit error rate increases with the increase of the number of antennas at the same signal to noise ratio, but the bit error rate tends to be stable with the increase of signal to noise ratio. With the increase of the signal to noise ratio, the bit error rate decreases. For the same antenna system, the BER of the antenna system adopting the algorithm is significantly lower than that of the antenna system without the same signal to noise ratio.

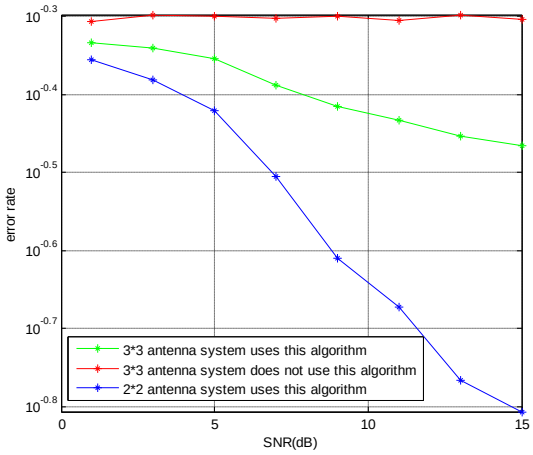


Figure 9. Error rate simulation diagram

It can be concluded that the algorithm can be used to simplify the receiving device of the receiver to a certain extent. But in practical applications, the interference of various signals in the channel also affects the signal of the receiver, so the algorithm needs to be improved so as to adapt to various channels.

VI. CONCLUSION AND PROSPECT

With the rapid development of science and technology, the demand for wireless service is not only the improvement of information rate, but also the high quality of receiving information. The space reuse and diversity technology of

MIMO multi antenna communication system can effectively transmit and receive multiple data streams at the same time and in the same frequency band. Reasonable use of spatial multiplexing and diversity technology can not only improve information rate, but also improve system performance. It can greatly improve the spectrum utilization of communication systems and meet the high rate of users' communication needs. Therefore, it has received extensive attention and research at home and abroad. It has become one of the most promising technologies in the 5G mobile communication system.

In the actual production and life, sometimes too many complex equipment will affect the results and feasibility of the experiment, so we should also pay attention to the study of the technology, and we should also pay attention to the simplification of the equipment without affecting the experimental results.

The purpose of this project is to bring forward a new principle based on the original spread spectrum, according to the theoretical knowledge, so that the equipment can be simplified. In this way, a higher effect will be achieved in the application of the 5G MIMO system. However, due to the interference of all kinds of noise in the signal transmission, it will have a certain influence on the receiving signal. Therefore, the algorithm can be improved from the direction of interference coordination to reduce the impact of interference on the received signal.

ACKNOWLEDGEMENTS

This project is supported by 'The Excellent Going Abroad Experts' Training Program in Hebei Province, Doctoral research fund of North China University of Science and Technology, National Science and technology support program (No. 2013BAH72B01), Hebei Natural Science Foundation of China: (No. F2014209276).

REFERENCES

[1] Guo Wenzhuo. Research on Key Technologies of multi antenna multiuser communication system [D]. Harbin: Harbin Engineering University, 2009.

[2] Zhang Fenghua, Li Wangqing. Application and development of spread spectrum communication technology [N]. Xinjiang science and technology daily (Han), 2005.

[3] Research on differential space-time coding technology in Chen Yuyan MIMO-OFDM system [D]. Shanghai: Shanghai Jiao Tong University.2006.

[4] GAN Y H, LING C MOW W H. Complex lattice reduction algorithm for low-complexity full-diversity MIMO detection[J]. IEEE Transactions on Signal Processing, 2009, 57(7): 2701-2710.

[5] ZU K, LAMARE R C, HAARDT M. Generalized design of low-complexity block diagonalization type precoding algorithms for multiuser MIMO systems[J]. IEEE Transactions on Communications, 2013, 61(2): 4232-4242 .

[6] Dong Tao, Jiang Zhucheng, you Xiao Hu. Optimization of channel estimation for spread spectrum system [J]. circuit and systems, 2004-02.

[7] F Adachi,M Sawahashi,H Suda. Wideband DS-CDMA for next-generation mobile communications systems[J].IEEE CommunicationsMagazine,1988,(09):56-69.

[8] Wang Huihua, Li Baoping. M sequence generator [J]. Journal of Beijing Electronics Science and Technology Institute, 2007,02:58-61.

[9] Liu Zhenming.Study on the cross correlation properties of m sequences [D].University of national defense science and technology, 2009.

[10] Cheng Jianli, Huang Fuqing, Tu Jia, Xu Yitao. The soft spread spectrum technology based on the same spread spectrum code [A]., the Youth Work Committee of the Chinese communications society, the new development of the.2007 communication theory and technology of the school of information engineering of North China University of Technology - the Twelfth National Youth communications academic association (book below) [C]. China Communications Society School of information engineering, North China University of Technology, 2007:4.

[11] Ma Yufeng, Lv Chengmin, Song Feng, Sun Jun. A research on the information hiding algorithm based on spread spectrum technology [A]. China Communications Society. The fifth academic annual conference of the China Communications Society, [C]. China Communications Society: 2008:5.[17] AI Bo. MIMO communication system coding [M]. Electronic Industry Press,.2008.

[12] Chen Hong.MIMO-OFDM key technology research [D]. Tianjin: Tianjin University, 2010.

[13] Nie Chunyan. Creative judgment of claims in communication network technology category [N]. China Intellectual Property Office, 2008.

[14] Ting Ann. Effective CRM software. Legacy systems can be accessed using integration technology.[J]. Healthcare Informatics, 2004, 21(2). [18]

[15] Joseph G. Young. Program analysis and transformation in mathematical programming[D]. Rice University, 2008.

[16] Comeron J M, Kreitman M. The correlation between synonymous and nonsynonymous substitutions in Drosophila: mutation, selection or relaxed constraints[J]. Genetics, 1998, 150(2).

[17] Tao Chongqiang, Yang Quan, Yuan Xiao. Simulation study on spread spectrum communication system of m sequence, Gold sequence and orthogonal Gold sequence [J]. electronic design engineering, 2012 (18): 148-150.

[18] Gu Jingmin, Liang Tao, Yu Yong. A new MIMO transmission power allocation algorithm research [A]. communication theory and new progress in signal processing - 2005 communication theory and signal processing annual conference paper, [C], 2005.

[19] Xiong Yan. Research on precoding technology of MIMO multi-user broadcast channel [D]. Beijing: Beijing University of Posts and Telecommunications.2011.

[20] Antenna selection technology in Fang Xiaoyong MIMO system [D]. Xi'an: Xi'an Electronic and Science University,.2009

.

The Design of Data Acquisition Based on the Spectrometer

Xu Shuping
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
e-mail：xusp686@163.com

Huang Mengyao, Xu Pei
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China

***Abstract*—Most of the spectrometer has the advantages of large volume, inconvenient carrying, slow data reading, long development cycle and high cost. In view of this situation, this paper presents a raspberry based portable spectrum analyzer. The general structure, working principle and data reading principle of the portable ultraviolet spectrum analyzer system based on raspberry faction are introduced in detail. And the raspberry sent a brief introduction, and finally introduced the raspberry pie to the development of the environment configuration, and the use of Python-Qt prepared control software for data read. The final test shows that the system interface is simple and friendly, stable operation, you can quickly read the data from the spectrometer, to meet the actual requirements.**

Keywords-Spectrometer; Raspberry Pi; Working Principle; Data Reading

I. INTRODUCTION

The micro-spectrometer has advantages of modular structure and flexible construction, so inpractice, only a spectrometer is needed, and the software can detect different samples in realtime [5]. At the same time, the micro-spectrometer has advantages of small size, easy to carry, compact internal structure, wide wavelength range, fast detection speed and low price. It has wide application development space in the field of industrial online monitoring and portable detection [8].

At present, the portable spectrometer in domestic and foreign market has two main methods: electrochemical analysis and infrared spectrometry[2]. The electrochemical analysis method has the advantages of simple structure and easy operation. It mainly depends on the gas sensor, and a gas sensor can only detect a corresponding gas, and the gas sensor also has a life limit, the use of a period of time, the sensitivity of the gas will be reduced, the need to replace gas sensors, and gas sensors are expensive, increasing the cost of use for the user. The main principle of the gas sensor is the use of gas oxidation or reduction reaction, the current, but if both the presence of oxidizing gas and reducing gas, the measurement results will be inaccurate [7]. Infrared spectroscopy overcomes the shortcomings of electrochemical analysis, but it can only measure the approximate concentration of NOx, and it cannot accurately measure the specific concentration of NO and NO2, and the infrared spectral analysis method is more complicated than the environment humidity, temperature, etc [6].

Based on the above problems, this paper presents a portable UV spectrometer developed using Maya2000Pro UV spectrometer and raspberry pi, according to the UV light wavelength 190nm-290nm, the gas has good absorption of ultraviolet light, and the measurement is absorbed by the gas, after the UV light, which can read through the spectrometer wavelength and photon number. The ultraviolet spectral analysis has the advantages of high reliability, strong anti-interference, and a variety of gases can be measured at the same time, accurate measurement and so on. The portable spectrometer has the advantages of fast data reading speed, low cost, convenient on-line debugging, stable operation and simple operation, and achieves practical requirements.

II. SYSTEM ARCHITECTURE

The portable UV spectrometer's system architecture as shown in Fig.1, and system consists of three modules: data acquisition module, data processing module and man-machine interaction module. The data acquisition module is composed of sampling probe, flue gas sampling pump, ultraviolet light source and micro spectrum analyzer, and the main purpose is to complete the collection of gas, analyze the spectral data before and after the ultraviolet light source and micro-spectrometer, and transform the spectral data into digital signal to transmit to the data processing module. In fact, in the whole system, the micro-spectrometer is a very important sensor component, and the spectrometer is mainly composed of optical components, photoelectric conversion module (CCD), memory. The optical components are mainly for optical fiber incoming optical signal processing, photoelectric conversion module (CCD) after the optical processing of ultraviolet light into electrical signals, memory used to store electrical signals [3]. At the core of the data center is the Raspberry Pi Development Board, then the task is to calculate the specific concentration of gases in the mixed gas by using the relevant algorithm. The man-machine interaction module is composed of a keyboard and an LCD screen, which is designed to visualize the concentration of gas and other related settings on line monitoring.

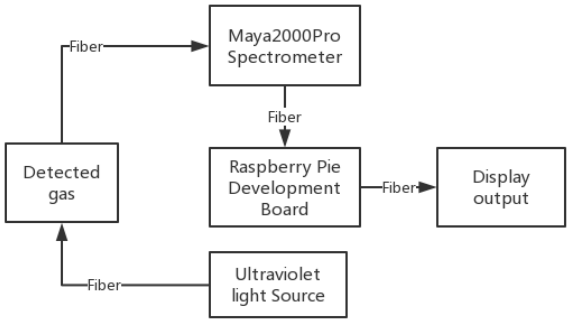


Figure 1. The overall architecture of the portable spectrometer

The key component of the portable UV spectrum analyzer control and data processing is the Raspberry Pi Development Board. It is a micro-computer motherboard based on ARM (shown in Fig.2), and the board contains an arm cortex A53 microprocessor, then the main frequency is 1.2GHz. It provides a 10m/100m Ethernet socket and four USB connectors, uses a normal SD card as a hard drive, supports C, C++, python and other development languages. The above parts are integrated in a motherboard, with the basic functions of ordinary computers, compared with the traditional embedded development platform such as ARM7, ARM9, Raspberry Pi in the development efficiency, operation speed, market price and so on have obvious advantages.

Figure 2. Raspberry Pi 3B Development Board

III. DATA ACQUISITION AND READING PRINCIPLE OF SPECTROMETER

A. Principle of data acquisition

The data acquisition of portable UV spectrometer is a process of collecting frames. When the spectrometer has collected a frame of data, will collect the spectral data into a queue, waiting for the Raspberry Pi sent instructions to read, and the Raspberry Pi reading spectral data is the process of USB or serial communication. The data read from the spectrometer is derived from the photoelectric conversion (CCD) module. The photoelectric conversion module used in the Maya 2000 pro series spectrometer is a thin photoelectric conversion (CCD) array on the back of the science level.

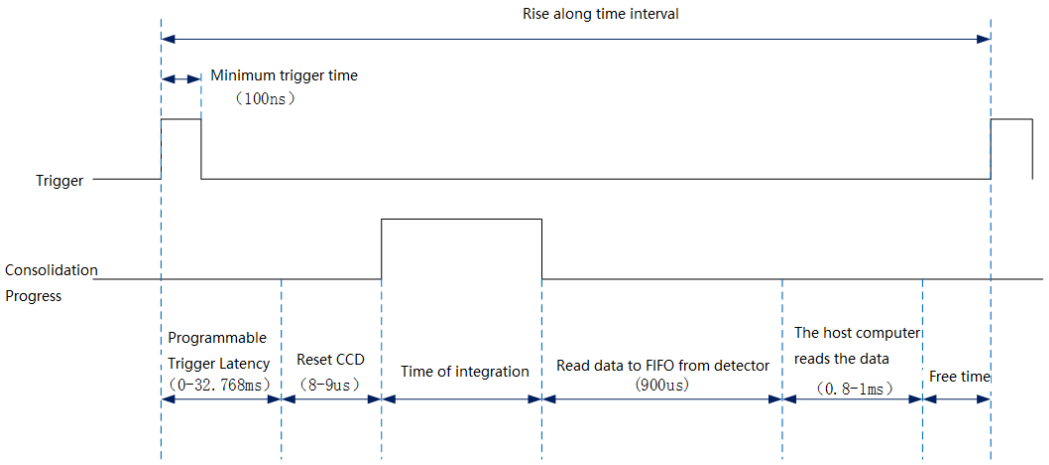


Figure 3. Data reading sequence diagram of spectrometer

The data reading sequence of UV spectrometer is shown in Figure 3. The spectrometer first receives the external triggering signal (high level, 100ns), at the same time, starts triggering the delay signal (low level), resets the CCD signal (low level signal), integration time signal (high level), and the data reads the signal (low level), Idle time signal (low level), after receiving the external trigger signal again, then a data acquisition process is over, and the collected data is stored in memory.

B. Principle of data reading

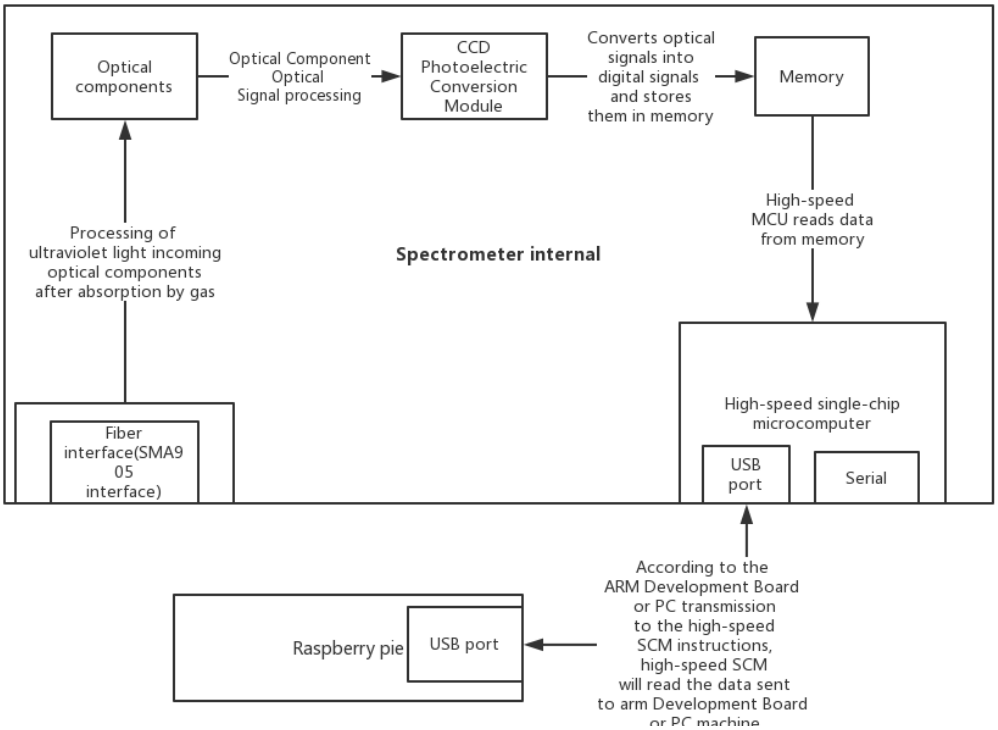


Figure 4. Schematic diagram of data reading of spectrometer

One of the most important components of the UV spectrometer is the high-speed single-chip microcomputer. High-speed single-chip computer has a USB port (square port, USB2.0, mini B-type 4-Pin) and serial port two interface (serial port for the 30-PIN Ocean Spectrometer special interface) [4].

The data reading process and principle of UV spectrometer are as follows:

First, the optical components of the ultraviolet light processing, ultraviolet light through the slit to the collimating mirror, after collimation, the grating is assembled onto the CCD after the grating is divided, and the photons with different wavelengths will hit the different CCD pixels after the light is divided, and then the CCD element converts the ultraviolet signal into the electrical signal. This completes the spectrometer and the detection function. The signal information obtained from the data will be stored in the memory, then, the raspberry pi is connected with the spectrometer through a USB port, and the raspberry and spectrometer communicate with each other through the USB interface. Finally, the raspberry sent to the spectrometer internal high-speed microcontroller to send instructions, and high-speed SCM will be based on instructions, from the memory to obtain relevant information, and return the information to the raspberry pi.

IV. RASPBERRY PIE DEVELOPMENT ENVIRONMENT CONFIGURATION AND SOFTWARE PROGRAMMING

A. Python-seabreeze Package Working principle

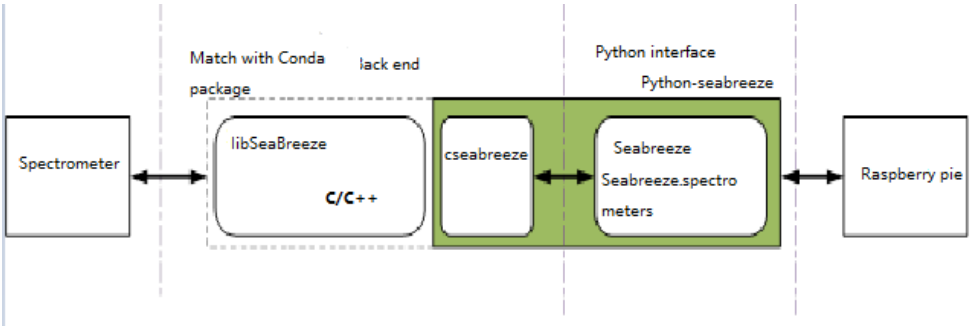


Figure 5. How Python reads the Data development package works

The Python-seabreeze package is a two-time development package developed by the Python language on the spectrometer and is the easiest way to access the Maya2000pro-type Ultraviolet spectrum analyzer in Python language. It communicates with the spectrometer using the Seabreeze Library provided by Ocean Optics. Any command sent by the raspberry pie to the ultraviolet spectrometer will use and rely on the Seabreeze library. Python-seabreeze is available as a prebuilt package on windows, Linux, macOS platforms in Python 2.7.x, 3.5.x, 3.6.x, and more. The package supports communication through the Cseabreeze backend with the Maya2000pro Series Ultraviolet spectrometer. In Figure 5, the Libseabreeze is a C + + language library ported from the Seabreeze 2.0 API. The back end is packaged for the Seabreeze Library and uses the Ocean Optics Open Source Seabreeze 2.0 API (using the "SeaBreezeAPI.h") interface. Almost all commands, such as the Format command for communication with the spectrometer, are completed in the C/S library. As shown in Figure 5, the Python language reads the data of the ultraviolet spectrum spectrometer through the Python programming interface provided by Python-seabreeze, and invokes Cseabreeze to read data from the ultraviolet spectrometer through the Libseabreeze library.

B. Raspberry PiDevelopment environment configuration (Raspberry Pi Development Board configuration environment must be networked)

1) Install the Python environment on Linux

The typical Linux system has a Python environment and is already configured. If not, you will install it by yourself. Python 2.7 is used in this article.

2) Install Python-seabreeze Package

Method One: In the Linux Terminal Command Window input command (raspberry pie must be networked, Python-seabreeze installation process is longer, please wait patiently): Conda install-c Poehlmann Python-seabreeze, The raspberry pie will then download the Python-seabreeze and install the configuration automatically.

Method Two: Download the Python-seabreeze installation package from the Internet, then copy the installation package to the installation directory, execute the python setup.py command in the command box, and install the configuration automatically.

After completion of 3.1.1 and 3.1.2, the installation environment is configured. Because of the use of raspberry pie, so here is the recommended installation Berryconda3-2.0.0-linux-armv7l,berryconda3 itself is a Python development package, and contains a large number of scientific calculation package, easy to install Third-party expansion pack (can use Conda for rapid upgrade), support Cross-platform operation, easy program migration, improve development efficiency, and this will also automatically install the modules and dependencies required for the Cseabreeze backend, saving a lot of time.

C. How Python reads data works

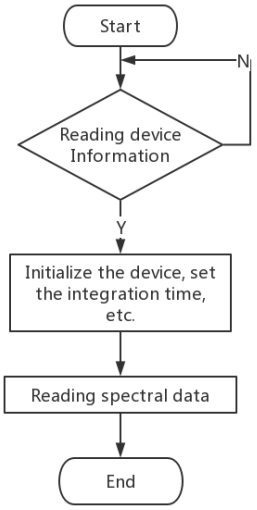


Figure 6. Data Reading flowchart

When the raspberry pie is reading the spectrometer data, it first reads the relevant information of the spectrometer equipment, and if it reads the equipment information of the spectrometer, initializes the spectrometer equipment, sets the integration time and so on; If you do not read the spectrometer device information, you will restart and read the spectrometer device information, cycle, until the device information is read. Second, after the spectrometer equipment initialization completes, the software starts to read the spectral data, and the basic data which reads from the spectrometer is the light wavelength and the photon number. This completes the spectrometer equipment data reads the work.

D. Raspberry Pie Data Read part of the code as follows

#Introduce related packages and device definitions

importseabreeze. spectrometers as sb#Introduction of Python-seabreeze Two development kits

devices=sb.list_devices() #Define spectrometer equipment

spec=sb. Spectrometer(devices[0])

#Device information

printspec.serial_number #Print output spectrometer factory serial number

```
printspec.model    #Print output spectrometer model
printspec.pixels   #Print output spectrometer number of pixels that can be read at a time

#Spectrometer Data acquisition
spec.integration_time_micros(10000)    #Set the spectrometer integration time
x=spec.wavelengths()
print x        #Print spectrometer wavelength data
y=spec.intensities()    #Print spectrometer photon Number data

#Drawing images using the readout wavelength and photon number
importmatplotlib.pyplot as plt
plt.plot(x,y,color="green",linewidth=2)
plt.xlabel("wavelengths")
plt.ylabel("integration")
plt.savefig("test2.png",dpi=120)
plt.show()
```

The drawing results are shown in the following illustration:

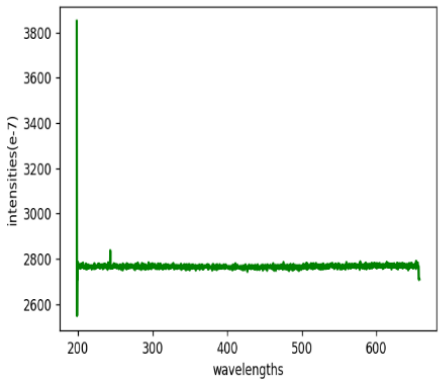


Figure 7. Wavelength of dark spectrum and photon number

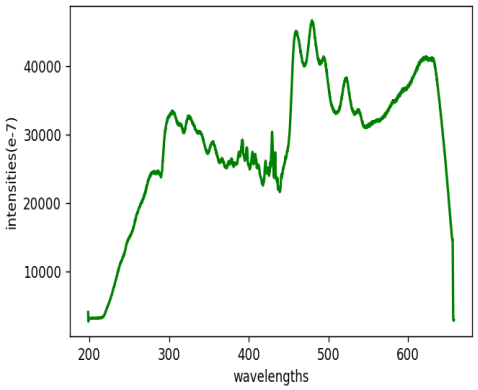


Figure 8. Wavelength of light spectrum and number of photons

Figures 7 and 8 are images drawn from the data (photon number and wavelength) that are read from the ultraviolet spectrometer. Fig. 7 is an image of the dark spectral wavelength and the number of photons, and fig. 8 is the wavelength of the light spectrum and the number of photons (the horizontal axis is the wavelength and the longitudinal axes are photon numbers).

Dark spectrum, which means that without an external light source, and the spectral line caused by dark current in the CCD of Ultraviolet spectrometer, when the ultraviolet spectrum is actually carried out, the number of photons measured by the detected gas needs to be reduced by the number of dark photons in order to get the actual photon number of the detected gas. It is shown from Fig.7 that the number of dark spectral photons is about 2,800, and the number of photons is related to the ambient temperature.

E. Software Simulation

The application is programmed with the Python Qt GUI. After writing, then the software is simulated and tested on the raspberry pie to verify that the software function meets the design standard, and the software simulation results are shown in Figure 9. Finally, the raspberry pie and spectrometer were debugged online (Fig.10).

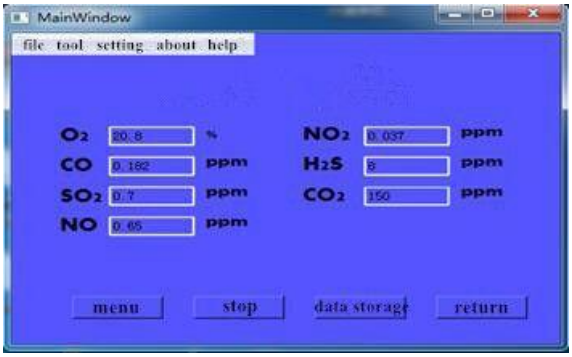


Figure 9. Software Run interface

Figure 10. Hardware and software online debugging

V. CONCLUSIONS

This paper mainly introduces the working principle of micro Ultraviolet spectrometer, data reading and so on, using the developed software after a long time test test shows: the theoretical analysis is correct, and the software has been written to achieve the raspberry pie from the micro-ultraviolet spectrometer read data, and work stable, fast, friendly interface, anti-interference ability, easy to operate and master. The system takes the raspberry pie as the core, realizes the data reading through two times development package, and carries on the software development through the python. The system has the advantages of short development cycle and stable operation, and achieves the practical goal. Next, based on this, the gas iterative algorithm is studied, and the actual concentration of various gases is calculated by using an iterative algorithm for the photon number read by the raspberry pie from the ultraviolet spectrometer.

VI. ACKNOWLEDGMENT

The authors wish to thank the cooperators. This research is partially funded by the Project funds in shanxi province department of education (15JF019)

REFERENCES

[1] Xu Shuping, Lichao. Portable smoke analyzer based on arm [J]. Computer measurement and control. 2011.19 (7): 1788-1790

[2] Lei Tianxue. Status of portable flue gas analyzer [J]. Environmental monitoring management and Technology, 1998, (4): 19.

[3] Shi Baosong, Sun Shouhong, Zhang Wei. Application of CCD in Portable Spectral analyzer [J]. Electronic measurement Technology, 2010 (11): 84-86

[4] Juwu ,WuYihui. Micro-miniaturization of spectrometer [J].Journal of Instrumentation, 2003. 22 (4): 131-133

[5] Yu Zhiqiang, Wenzhi Yu, XieYingke, Zhou Suyi. The control system of multi-parameter water quality tester based on raspberry pie [J]. Instrumentation technology and Sensors, 2015 (06): 20-23.

[6] Han Xiao, Wenzhi Yu, XieYingke, Wei Kanglin, Zhou Xiaofeng. Software design of control and signal processing system for multi-parameter water quality tester [J]. Instrumentation technology and Sensors, 2014 (08): 20-22

[7] Lichao. Design of a portable Smoke analyzer controller based on ARM [D]. Xian: Xi ' an University of Technology, 2012

[8] Zhang Yi. Design of portable multifunctional flue gas analyzer [D]. Xian: Xi ' an University of Technology, 2012.

A CEP Privacy Security Access Control Framework Based on Event Attribute Detecting Tree

Bo Hong*
School of Computer Science and Engineering,
Xi'an Technological University,
Xi'an, 710021, China
e-mail: 1964383053@qq.com

Xin Jing
School of Computer Science and Engineering,
Xi'an Technological University,
Xi'an, 710021, China

***Abstract*—Complex event processing (CEP) technology is a study focus in the data flow processing area, while privacy security protection is the key problem that needs to be solved. In order to prevent illegal users from acquiring any information via registered event patterns, this paper discusses the CEP privacy security access control object in depth, formally defines four types of event attribute operators including completely read, partially read, access denied and quantity statistics, presents a privacy security protection engine with the event attribute detecting tree as the operating mechanism and puts forward a new feasible CEP privacy security access control framework based on this. The experimental result shows that such framework is able to realize efficient privacy information filtration based on the user role to reach the goal of CEP detecting information processing in a safe manner.**

Keywords-Complex Event Processing; Privacy Security Access Control; Event Attribute Detecting Tree; Event Attribute Operator; Security Protection Engine

I. INTRODUCTION

Data flow processing is a very important and active area in modern database technology. CEP technology[1] has become the study focus of such field since its inception as it is capable of integrating the information from the numerous data source distributed and digging the valuable dynamic meaning among the information from the high-speed data flow in real time. CEP technology is thoroughly changing the way of subscription & distribution and application data of the traditional information system. It acts as the hub of information fusion and dispersion by uncoupling the information provider and recipient and playing the roles including information observer, analyst and decision maker. As the Internet of Things sensor and the network based new application quantity surge, the information capacity to be processed sees an explosive growth trend. Thus, CEP technology is increasingly becoming an essential tool in many application fields. However, for most CEP engines at present, the processes and content of the complex event processing and output are open. That is to say, not only legal advanced application can utilize the CEP engine to obtain valuable information, but also illegal users are also able to acquire any necessary information for their criminal behaviors. This presents the CEP technology with huge responsibility with respect to the privacy security protection in detecting information.

Up to now, there are few studies on CEP privacy security access control, thus the research result in such aspect is just in the initial stage. In order to hold back over-class information access, literature [2] conducts security access expansion for the CEP detection and event model, which effectively prevents the unauthorized information from being leaked or tampered to the outside. It first increases two attribute fields, i.e. "security level" and "current stage", behind the traditional event model, and then adds security level checker in the query matching tree. The checker allows the event the security level of which is lower than the level set by this query matching tree to inflow so as to realize access control of the information at different security levels. Literature[3, 4]designs a set of novel security access operators and comes up with a re-query method based on such operator set with the relation algebra and query graph model of Aurora as well as the view idea of the traditional

database management system. Through this method, the security access operators are able to be inserted to the Aurora query graph model in the most effective way. As a result, CEP can perform security access control on the data flow in pursuant to the predefined security strategy file. The above studies share the same thought, i.e. rewrite the query of the CEP event pattern, adjust the original performing structure, and insert specialized security detecting unit to form an operation structure combining security and the original detection pattern. Such kind of method is complex and has some deficiencies. 1. Users have different security strategies, thus it is necessary to save all relevant user security strategies in the security detection unit when performing multiple user security strategies in one CEP query, which obviously will cause logical mess in the course of performance. 2. The newly added security detection unit will produce more work load in the process of CEP detection and influences its execution efficiency, meanwhile the mixed operation structure will be hard to be optimized (e.g. share intermediate result). In order to avoid the above problems, this article will put forward an efficient CEP privacy security access control framework that is feasible and easy to be integrated.

II. CEP PRIVACY SECURITY ACCESS CONTROL OBJECT

The basic unit of CEP processing work is event. Thus, the content of its privacy security access control is the information included in the event. According to the event model definition provided by the author in the early stage of the study (Event_Model:= Event_Type @ (Attribute_Name[Data_Type]n) n≥1;), event is a tuple composed of N attributes (A1,…,An) and attribute field is the minimum unit saved by the information value. Therefore, this paper determines event attribute as the object of CEP privacy security access control and explain its concept in the form of definition.

Definition 1 Event Attribute It specifies that each event flow Stri input into the CEP engine contains one type and can only contain one type of event ETj. Certain event type ETj is made of N attributes Pk k≥1. P(Stri) represents the set of all attributes included in certain event flow Stri and P(ETj) represents the set of all attributes included in certain event type ETj, then P(Stri)=P(ETj) if ETj∈Stri. In addition, in this paper, Stri.pk (or ETj.pk) represents certain attribute in certain event flow (or certain event type).

The user's access right to the event attribute (ETj.pk|Stri.pk) content meets four cases: Completely read, partially read, access denied and quantity statistics. Therefore, a formalized description of such four types of access control operator is firstly given.

Completely read operator ξ: ξ(P(ETi))|ξ(P(Stri)) represents complete access control right to the event attribute information in the event type (or event flow). It can be abbreviated as ξ(ETi)|ξ(Stri). ξ can be used for some attributes set of the event type (or event flow), ξ(ETi[p1,p2,…]) p1,p2,…∈P(ETi) means only the information content of some attributes (p1,p2,…) in the event type is allowed to be accessed.

Partially read operator ϕ: ϕ(Expr)(ETi)|ϕ(Expr)(Stri) means the event attribute information in the event type (or event flow) can be accessed as per the definition of the conditional expression set Expr. The expression expri in Expr expression set only exists as conjunction relationship, e.g. ETi.location="L1"∧ ETi. temperature>30, means the location attribute of such event is L1, and the temperature value attribute is greater than 30.

Access denied operator ψ: ψ(P(ETi))|ψ(P(Stri)) represents complete denial of the access to the event attribute information in the event type (or event flow). It can be abbreviated as ψ(ETi)|ψ(Stri). Likewise, operator ψ can also only deny the access to some attributes, ψ(ETi[p1,p2,…]) p1,p2,…∈P(ETi) means only the information content of some attributes (p1,p2,…) in the event type is denied to be accessed.

Quantity statistics operator Ω: This access operator corresponds to aggregate operations that do not care the specific value of the event attribute but concern the total number, mean value and other statistics information of the event. Ω(F(Pk))(ETi)|Ω(F(Pk))(Stri) means it has statistical right to the event attribute Pk in the event type (or event flow), of which, F is calculation function, including min, max, count, avg and sum, etc.

In pursuant to the above formalized description of the control operators of security access to event attribute (record all operators set Э), the content to which the user may have privacy security access for the CEP input event flow should substantially be the result of Э operation on the input event flow by such user. That is to say, only the information in line with the given user security strategy is filtrated. Based on this, this article defines CEP privacy security access control object as follows.

Definition 2 Privacy Security Access Control Object The privacy security access control object in CEP engine is, of which, Strs is the input event flow set of the CEP engine, Э is the set of the security access control operators of event attribute, and Pi is the event attribute set in the event flow.

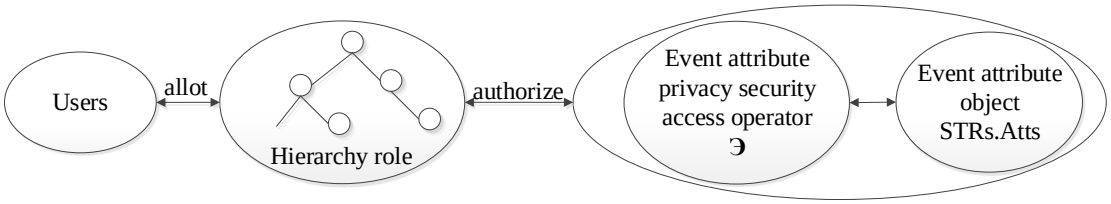


Figure 1. Privacy security access control model

As shown from the above definition, when allocating security access control right to a user, the system administrator needs to explicitly designate the privacy security protection object to which such user can access for such user, i.e. designate the access right to each event attribute content for such user. This will bring huge work load for the system administrator. In order to operate flexibly and conveniently as well as reduce the work load on right allocation, this paper divides the security access control right of users based on the RBAC model and hierarchy role thought. As shown in Fig.1, hierarchy role applies tree structure. High level role may include several predecessor roles and will automatically inherit the security access rights of all predecessor roles to event attribute. Similarly, a user instance may have one or more role identities so as to realize flexible role allocation.

III. CEP Privacy Security Access Control Frame

According to the above privacy security access control object, this paper presents CEP security access control framework (CEP-SACF) as shown in Fig.2. CEP-SACF is easy to be realized without changing the original CEP implementation structure.

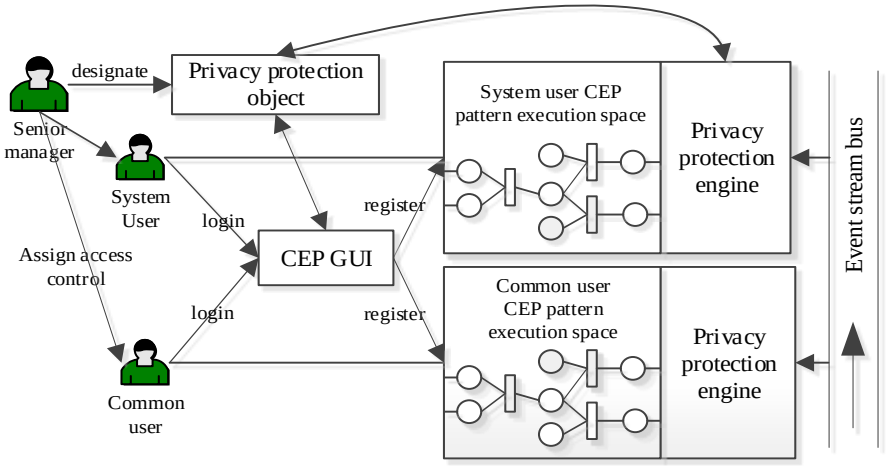


Figure 2. CEP privacy security access control framework

CEP-SACF operation includes three stages, i.e. user role authorization, pattern registration and privacy security access control. First of all, at the user role authorization stage, senior system administrator defines the privacy security access control object, designates corresponding security access operation to the event attribute requiring privacy protection and allocates it to the designated user role (user role is planned by level,[5, 6]to reduce authorization work load); then, the user logs in with the role allocated and enters CEP engine management interface where user may define its own business rule with CEP event pattern language[7], CEP manages GUI and will correlate the security rules related to such user role in the privacy security access control object strategy file to detect the legality of the event pattern to be registered. If there is no conflict of security rules, then such event pattern will be registered in the independent implementation space of such user role (at the time of the first registration, an independent operation space should be firstly created for such user role, and the event pattern hereafter will be registered under the namespace with the same name as the registered user role). If the event attribute content requested by the event pattern to be accessed is in conflict with the privacy security rules of this role, then a prompt of limited user right will show and registration of such event pattern will be denied; finally, the independent privacy security protection engine will operate between the input event flow and CEP engine. Each PSPE just saves the privacy security rules related to such user role and will make operations of permission (completely read operator ξ), rejection (access denied operator ψ), filtration (partially read operator ϕ) or modification (quantity statistics operator Ω) for the event attribute in accordance with the definition of privacy security access operator of event attribute $\Im$. The event processed by operator Ω will be repacked. For example, certain event contains (ID, TimeStamp, Location) attribute previously and such event only permits calculate the total number (perform Ω operation for its ID attribute). Other attributes are private information that is not permitted to be accessed. Then under the function of operator Ω, PSPE will allow all such events to pass with the private information contained flowing through the event removed. A new event only containing ID attribute will be generated. Then it will be sent to the corresponding operation space. Furthermore, in order to ensure security of CEP output result, PSPE will also receive the output in the operation space protected by it and send the result to the user within the user role of such space.

To ensure that under the registered event pattern, the user will not acquire the privacy security access right designated to such user role beyond the senior administrator and guarantee the efficiency of legal detection, this article verify the event pattern registered by the user with the following algorithm.

Algorithm 1 Validity Verification Algorithm of CEP Privacy Security Access Control in Event Pattern

Input: The event pattern declared by the user and user role;
Output: The event attribute array NProps[] without legal access right in the event pattern definition;

```
1. if (find Prop.aggregation(*) in Event_pattern)==true
      Props<String,String>.put(EventType,aggregation_operator_name);
2. Iterator (expression in where clause ) {
      EventType=get_EventType(in expression);
      Property=get_ Property(in expression);
Props<String,String>.put(EventType, Property);}
3. for(Map.Entry<String, String> entry:Props.entrySet()){
      Select * from Secunity_rule where user_role=login_user_role;
for( Dataset.hasNext() ){
if (Dataset[i].eventProperty==entry.getKey())
if (NoLegality(Dataset[i].accessOperator,entry.getValue())==ture)
          NProps[entry.getValue()];}}
4. System.out.println(NProps[]);
```

IV. PRIVACY SECURITY PROTECTION ENGINE (PSPE)

PSPE is independently created with CEP event pattern implementation space. Its content is determined by the privacy security access control rules defined by the senior system administrator and automatically updated depending on the adjustment of the rules. The basic mechanism of PSPE operation is copy, i.e. regard the event flow input into CEP engine as data bus and send the copy of the event in line with the privacy security protection rules on the data bus to the event pattern detection network inside the operation space. Beyond that, no operation will be made. This mechanism can effectively guarantee the event flow will flow through all privacy security protection engines and finally pass the event containing correctly authorized information to CEP processing nodes.

The working principle inside PSPE is shown in Fig.3. It will convert the filtration operation of the event attribute to the tree structure with the event type as the root node, of which, EventType is the event type that can be processed in this space. The subnode under the root node of the event type is the event type included. The event attribute node will be included in the access operation defined in the privacy security protection rules (as one event attribute can only define one type of security access operation type, the event attribute node only contains one subnode).

Here are some kinds of common detecting tree in PSPE. As shown in Fig.3 (a), suppose certain event type contains three event attributes and for certain user role, these three event attributes are all permitted to be accessed, thus the combined node will pass such event to the internal implementation space completely. It is contrary in Fig.3 (b) where the three attributes of the event are denied to be accessed, and such event will not be passed internally. Fig.3 (c) shows the general situation under privacy security access control, i.e. user role is only allowed to access to some attribute content of one event while the private part is not permitted to be viewed. As Attr2 attribute is denied to be accessed, the node of such detecting tree will only combine Attr1 and Attr3 attributes and outputs a new event which only contains these two attributes. Fig.3 (d) displays the appearance of the detecting tree which conditionally reads the event attribute, of which, the condition verification includes single value comparison (as shown in Figure 2 (e), the comparison content: Attr1= value 1 && Attr3!= value 2) and multiple value comparison (as shown in Fig.3 (d), the comparison content: value 2<Attr2< value 1). The node will only allow the event whose comparison result is true to pass through. Fig.3 (f) shows the situation of event attribute statistics and calculation. The node will permit such event attribute content to be accessed and the function of node Ω is equivalent to ξ. As known from the above common detecting tree structure, PSPE is able to effectively prevent the unauthorized information from inflowing and using. By means of repackaging the event, the separation of the authorized and unauthorized information can be guaranteed.

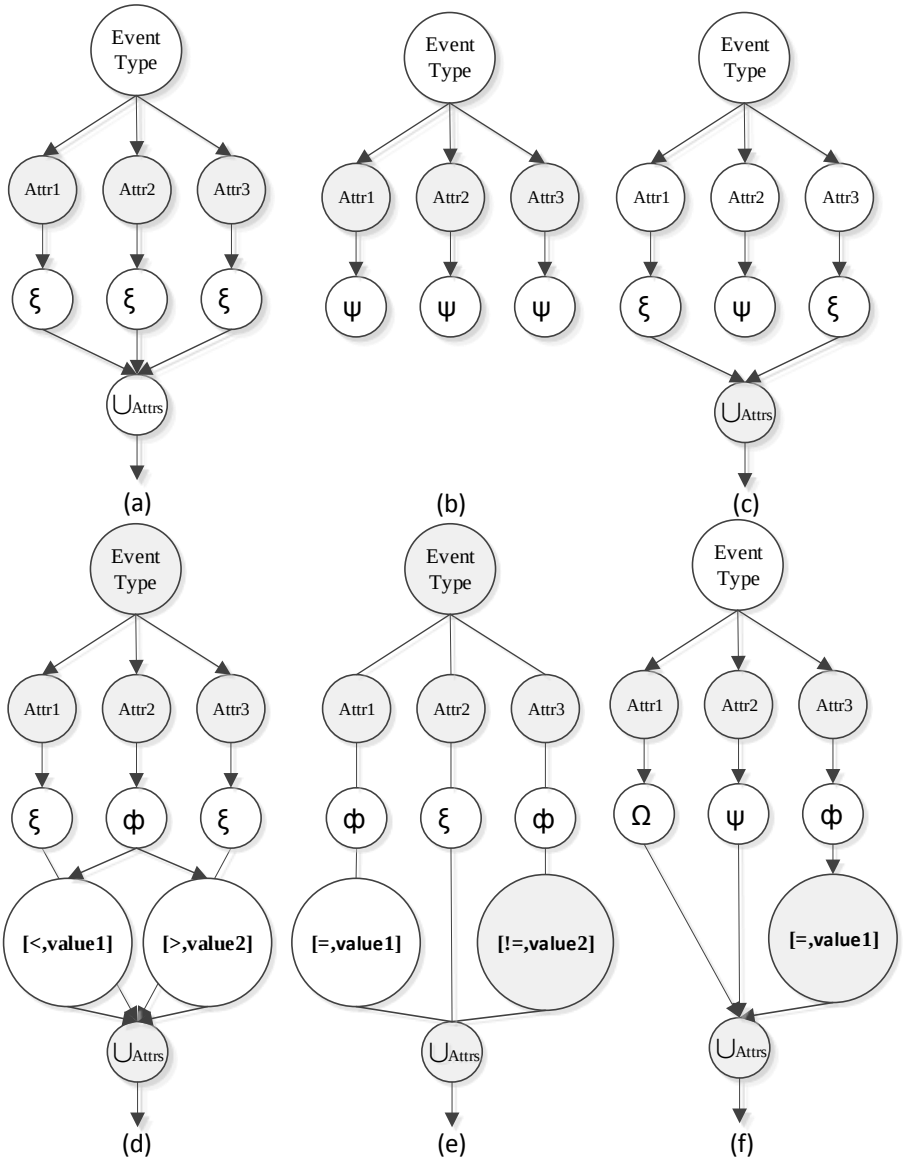
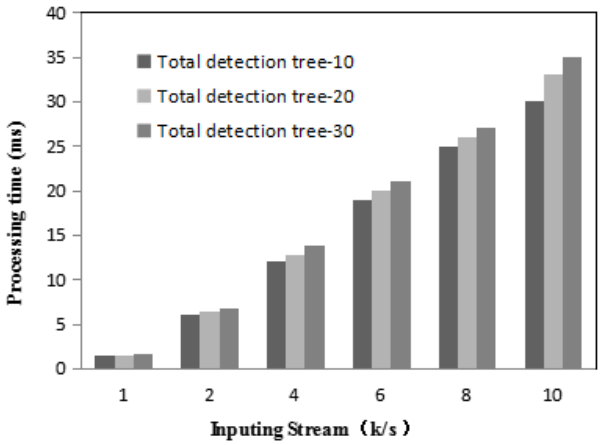


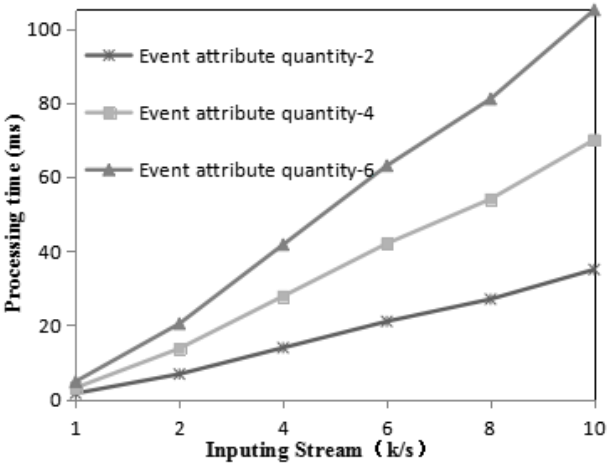
Figure 3. Detection tree structure in privacy security protection engine

V. Performance Test of Privacy Security Access Control Framework

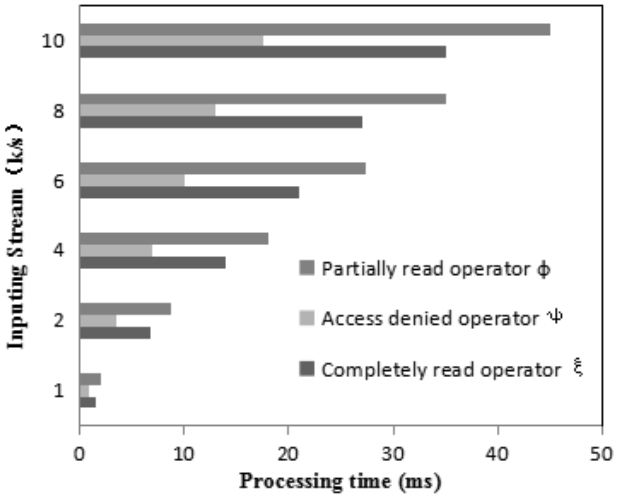
The core part of CEP-SACF operation is PSPE, the operation performance of which is related to the input event flow rate and the total quantity of the internally registered detecting tree (record such parameter as ETs). The above content shows that the working efficiency of the detecting tree is related to the number of internal event attribute node (record such parameter as ATs) and the node type of the access operator (record such parameter as OPs). Therefore, this group of experiment will test the three parameters that influence the efficiency of the engine respectively. First of all, simulate the input event flow, each of which only contains one type of event. Each event is composed of one event type attribute field and several other attribute fields. The event flow generator will utilize multiple courses to produce event flow in parallel and send it out to simulate real scene. Then, the buffer queue of PSPE will receive these events and conduct security detection by the first-in-and-first-out sequence. The detecting tree indexes with the hash table and realizes it with the custom tree structure.



(a) Experiment 1 influence of detection tree number



(b) Experiment 2 influence of event attribute number



(c) Event attribute security access operator efficiency test

Figure 4. Privacy security protection engine performance testing

Experiment I registers three groups of Ets={10,20,30} with different quantity for contrast. It receives data at 1,000-10,000 events/s and provides that the attribute quantity included in all events of such group of experiment is 2. The event attribute access operator is completely read operator ξ. As known from Fig.4 (a), as the input rate increases, the time taken by PSPE presents a linear growth trend, but the rise of total Ets has little influence on PSPE implementation efficiency because Hash Index has a high efficiency. The growth of total Ets has little influence on its index rate.

Experiment II tests the three control groups in which the attribute quantity of the event is Ats={2,4,6}. The total detecting tree registered in such group of experiment is Ets=30 (thus, the actual total number of the attribute detecting tree in PSPE is 60, 120 and 180, respectively). Also, it receives data at 1,000-10,000 events/s and provides that the event attribute access operator of all attributes is completely read operator ξ. As known from Fig.4 (b), parameter Ats has a great influence on the implementation efficiency of PSPE. As the total Ats increases, the calculated amount of the traversal node inside PSPE will undergo a cumulative rise. The processing time taken by the three control groups basically keeps a multiple relationship. The total consuming time of PSPE is at millisecond level, which has little influence on the overall operation efficiency of CEP-SACF.

Experiment III tests the performance of ξ, ϕ and ψ (as Ω and ξ is different in function, repeated test will not be done for Ω). This group of experiment provides Ets=30, Ats=2, Ops={ξ, ϕ, ψ}, with the data flow rate the same as above. The conditional expression of ϕ is [>,0]. That is to say, in spite of conditional judgment, all events are permitted to pass through. According to Fig.4 (c), as ψ denies events to pass through and there is no subsequent treatment. Thus, it consumes the shortest time (only including the time consumed in event type node searching and event attribute transversing). ϕ has calculation of conditional judgment on its node, so it consumes more time than the benchmark ξ operation. However, they come to the same conclusion that for different operators at different input rate, the total time

consumed in processing by PSPE can still keep at millisecond level, which represents high operation efficiency.

VI. CONCLUSION

The experimental result shows that PSPE has a high implementation efficiency. At the same time, it is able to deal with different user roles in different ways, filtrate the event information not allowed to be accessed and generate new events in line with privacy security access control requirement. Thus, it has a feature of customizability. In addition, PSPE is completely integrated outside of CEP engine, which is very feasible because it has no influence on its original implementation structure and operation efficiency. It has certain application value by effectively making privacy security detection on the event information input/output CEP engine.

ACKNOWLEDGMENT

This work was supported by the Industrial research project of Science and Technology Department and the office of Education of Shaanxi Province(Grant No. 2016KTZDGY4-09, 17JZ004), and the Xi'an Technological University Principal Fund (Grant No. XAGDXJJ15015).

REFERENCES

[1] Luckham D. The power of events: An introduction to complex event processing in distributed enterprise systems[M]. Springer, 2008.

[2] Buddhika T, Ray I, Linderman M, Jayasumana A. Secure complex event processing in a heterogeneous and dynamic network[C]. SPIE Defense+ Security, International Society for Optics and Photonics,2014:907907-907913.

[3] Carminati B, Ferrari E, Tan K L. Enforcing access control over data streams[C]. Proceedings of the 12th ACM symposium on Access control models and technologies, ACM,2007:21-30.

[4] Carminati B, Ferrari E, Cao J, Tan K L. A framework to enforce access control over data streams[J]. ACM Transactions on Information and System Security (TISSEC), 2010,13(3):28.

[5] Tang Jin-peng, Li Ling-lin, Yang Lu-ming. User attributes oriented RBAC model[J]. Computer Engineering and Design, 2010,(10):2184-2186.

[6] Xiong Hou-ren, Chen Xing-yuan, Zhang Bin, Yang Yan. Security Principles for RBAC-based Authorization Management[J]. Computer Science, 2015,42(3):117-123.

[7] Jing Xin, Zhang Jing. Research on Parallel CEP Processing with the Multi-Event Pattern Shareing Capability[J]. Journal of Xi' an Technological University, 2014,34(9):715-719.

The New Method of Sea-sky Line Detection Based on Mathematical Morphology

Zhang Wenqi
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
e-mail:362976306@qq.com

Yu Jun
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
e-mail: 763757335 @qq.com

Bai Wanmin
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
e-mail: 541592039@qq.com

Gao Shouyi
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
e-mail: 478204287@qq.com

***Abstract*—To solve the problem of low accuracy and robustness of sea-sky line detection, this paper presents a method of sea-sky line detection based on the mathematical morphology. Firstly, the mathematical morphology closed-open operation is used to filter and denoise the sea-sky image. Then the Canny operator is used to obtain the sea-sky boundary of the image, Then mathematical morphological operation is used to remove some disturbing points. Finally, the sea-sky line is detected by Hough transform. The experimental results show that the algorithm can accurately and efficiently detect the sea-sky line under the complex sea-sky background.**

Keywords-Sea-sky Line; Mathematical Morphology; Edge Detection; Hough Transformation

I. INTRODUCTION

The sea-sky line is the dividing line between the sea and the sky. In general, an image of sea-sky background mainly includes three regions, those are brighter sky area, the darker sea area, and the sea-sky line area from light to dark [1]. If the low altitude investigation is carried out, the target on the sea usually appears in the area of sea-sky line. By detecting and obtaining the sea-sky line, it can reduce the calculation amount of target detection and shorten the calculation time. At the same time, it can distinguish between the sky area and sea area, that is meaningful to the simulation experiment of target detection on the sea.

The sea-sky line detection is influenced from marine environment greatly. The main influencing factors are as follows :

(1) The strong watermark interference caused by the wave, that makes the gray-value of the wave which is close to the pixel point gray-value of the sea-sky line, so that the extraction of the sea-sky line is difficult.

(2) When the background images contain mountains, ships and so on , which will interfere with the detection of sea-sky lines;

(3) When the atmospheric visibility is low, the sea-sky boundary is blurred, which leads to difficult y on detecting sea-sky line.

In order to detect the sea-sky line accurately, we need to know about its characteristics as follows.

(1) The area of sea-sky line is between the sky and the sea. Its brightness is more intense than the other two parts. Grayscale changes strongly in vertical direction as well as varies in horizontal direction slowly.

(2) The sea-sky line is usually not a straight line but a gradual change band.

At present, there are many reference documentation on sea-sky line detection. For example, Liang D and others use the algorithm of OTSU segmentation and clustering in order to detect the sea-sky lines[2]. Because the OTSU segmentation algorithm could not accurately segment the sea-sky background images with imbalanced illumination. It makes the sea-sky line detection error in this kind of image; H. Wang and others use the algorithm of combining the Sobel operator with the straight line fitting to carry out the sea-sky line detection[3]. This method can be used to extract the sea-sky line in simple background, but it is difficult to get a satisfactory extraction effect in some complicated situations. Wang Bo and others use the algorithm of gradient saliency region growth to detect the sea-sky lines[4], but the sea surface splash and water wave will interfere with the image gradient calculation. For the complex images of sea conditions, these methods are limited in certain degree.

In order to improve the robustness and accuracy of sea-sky line detection, the method of sea-sky line detection based on mathematical morphology is proposed. This method can improve the robustness of sea-sky line detection in the complex sea-sky background. The mathematical morphology are used to denoise sea-sky images and remove interference points, which can reduce computation and improve the accuracy of sea sky detection.

II. MATHEMATICAL MORPHOLOGY

The mathematical morphology is a nonlinear image processing and analysis theory. It is characterized by geometrical method, which is more suitable for the processing and analysis of visual information. The basic idea of the mathematical morphology is to measure the availability of the target image region and the effectiveness of the filling method by using a certain form of structural elements. Then it extracts more essential information of the related characteristics of the image morphology ,which can achieve the purpose of the target image analysis and recognition. The mathematical morphology can eliminate the unrelated morphological and structural attributes in the target image and retain the basic nature of the morphological and structural properties to simplify the target image data, so that it has the characteristics of fast parallel speed and easy implement in hardware. The algorithm has the natural parallel structure. It realizes the parallel of morphological analysis and process, which greatly improves the speed of image analysis and process.

At present, the mathematical morphology has been widely used in the fields of pattern recognition, machine vision, microscopic image analysis, medical image processing, computing and data processing and so on. It has obvious advantages in image processing problems such as filtering noise reduction, image enhancement, edge detection, image segmentation, feature extraction, texture analysis, image restoration and reconstruction, and image compression and so on.

A. The.Mathematical Morphology Operation

The mathematical morphology is composed of a set of morphological algebraic operators , whose basic operations are shown as follows: expansion, erosion, opening and closing. These operations have different characteristics in binary and grayscale images [5] . These basic operations can also be derived and combined into various practical algorithms of the mathematical morphology, which can be used to analyze and process the shape and structure of the image.

The most basic morphological transformation of the mathematical morphology includes expansion and corrosion, which can achieve many functions, such as filtering noise, dividing the independent elements and bridging the adjacent elements in the target image. The mathematical morphology can also be used to find the maximum or minimum region of the obvious block in the target image and get the gradient of the target image.

The expansion operation is to calculate the local maximum. While the corrosion operation is to calculate the minimum value of the pixel in the area. The two operations are a pair of mutually dual operations [6].The expansion operation is a process to expand the edge to the outside. It can be used to fill the small holes in the target image and transform the background edge into the target edge, so that the goal is increased and the background is reduced. The corrosion operation is a process that removes the unrelated edge points and makes the edge shrink inward. It can eliminate the small bulges in the target image and reduce the target and increase the background.

The other morphological operations are composed of two basic morphological transformations[6-7], such as opening and closing. The f(x, y) is set as an input image, b(x, y)is set as a structural element. The structure element b is used to handle the input the image f(x, y).As the formula(1) shown, this is an expansion operation. As the formula (2) shown, this is a corrosion operation .

Definition 1 The image f is expanded by using the structural element b , writed as $f \oplus b$.

$$[f \oplus b](x, y) = \max_{(s,t)\in b}\{f(x - s, y - t)\} \tag{1}$$

Definition 2 The image f is Corroded by using the structural element b , writed as $f \ominus b$.

$$[f \ominus b](x, y) = \min_{(s,t)\in b}\{f(x + s, y + t)\} \tag{2}$$

Based on the two basic morphological transformations of expansion and corrosion, many mathematical morphological clusters can be constructed. While open and closed operations are the two basic operations in the cluster [7-8] . The open operation firstly corrodes the image and then expands it, as shown in the formula (3).The closed operation firstly expands the image and then corrodes it, as shown in the formula (4).

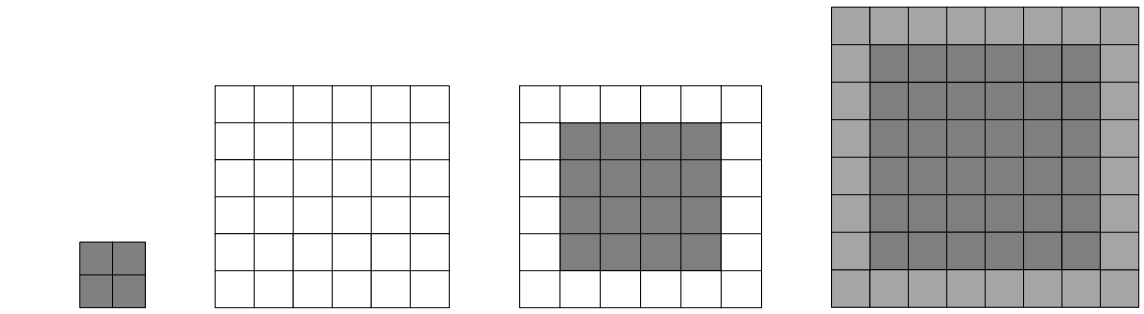
Definition 3 The image f is opened by using the structural element b , writed as $f \circ b$

$$f \circ b = (f \ominus b) \oplus b \tag{3}$$

Definition 4 The image f is closed by using the structural element b, writed as $f \bullet b$

$$f \bullet b = (f \oplus b) \ominus b \tag{4}$$

As shown in Fig.1, the Fig.(a) is a square structural element, whose size is 2×2. The Fig.(b) is the objective matrix that we want to perform the mathematical morphology operations. The Fig.(c) is the result diagram of the corrosion operation. We use the structural elements that is shown in the Fig.(a) to handle the target matrix of the Fig.(b). The Fig.(d) is the result graph of the expansion operation that we use the structural elements that is shown in the Fig.(a) to handle the target matrix of the Fig.(b).



（a）Structural elements （b）Matrix （c）Corrosion operation （d）Expansion operation

Figure 1. The mathematical morphology operation

The open operation can remove isolated points, burrs and small bridges (that is, the small points connected to two blocks), which can be used to segment large areas and smooth the edges of large area. While the total position and shape are constant. The closed operation can fill the small holes in the object and achieve the purpose of stitching small cracks to connect the adjacent objects and smooth edges .While the total position and shape are constant. The open operation and closed operation are also a pair of dual operations.

The closed operation can be filled with low grayscale black holes. And the open operation will inhibit the white point(noise) with high gray value. The operation that the closed operation is performed firstly ,followed that the open operation is used to make the de-noising effect better and smooth edge. Therefore, the closed - open operation are choosen in the this article to filter and reduce the noise of the sky-sea images.

B. Selection of Structural Elements

In any condition, the mathematical morphology algorithm is composed of two basic problems: mathematical morphology operation and structural element selection. The definition of mathematical morphology makes the operation rules of mathematical morphology constant. Therefore, the selection of morphological and structural elements determines the purpose and effect of mathematical morphology algorithm. Throughout, The determination and optimization of structural elements have become hot topics and difficulties in the study of the mathematical morphology.

The choice of the morphological structure elements can be divided into two aspects: the size and the shape of the structural elements. Generally speaking, the structural elements must be geometrically simpler than the original image. And they are bounded; Besides, the convexity of structural elements is also important. Based on the selection principle of structural elements, we usually choose some small simple collections, such as square, diamond, circle and so on. As shown in the Fig. 2, there are some examples of structural elements.

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

（b）The linear structural elements （b）The rhombic structure element

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

(c)The rectangular structural elements （d）The square structural elements

Figure 2. The structural elements

If the structural elements are not properly selected, they can not effectively process pictures. And the result will not have the desired effect. That mainly include the following two conditions.(1) When the size of the selected structure element is too small, the open operation cannot effectively eliminate the larger high grayscale noise point; For larger, low-gray black holes, that cannot be effectively bridged by the closed calculations. (2) When the size of the selected structure element is too large, on the one hand, the open operations will excessively eliminate pixel points on the edges of the image and cause false breaks. On the other hand, the closed operations will over-combine black holes and generate interference information [9-10] .

Therefore, the use of single size structure elements can easily lead to the edge location of the target image is not accurate enough and the denoising effect is not ideal. In

addition, due to the existence of a constraint relation on the edge of the image, the noise of the image is generated randomly.

When the structural elements is used to measure the target image, a geometric shape similar edge point can always be found near the edge points of the image. Thus, it is not effective to retain the edge segmentation information of the image by using a single morphological structure element to extract the edge of the target image.

Consequently, when the sea-sky line is extracted in the target image. If the single size and shape structure elements are used for image processing, the location is not accurate, the de-noising effect is not ideal, and the detail information of the sea-sky line can not be retained effectively. In conclusion, this paper adopts multi-dimensional, multiple-shape structure elements to process sea-sky images.

III. SEA-SKY LINE DETECTION ALGORITHM DESIGN

The sea-sky line detection algorithm in this paper is based on the mathematical morphology. The overall flow diagram is shown in the Fig.3 . Firstly, the sky-sea image is preprocessed. The mathematical morphology is used to filter the target sea-sky background image and denoise the interference of the sea-sky lines; Secondly, the Canny operator is used to extract the sea sky boundary of the preprocessed sea-sky background image. Follow that the mathematical morphology is used again to remove the interference points, so that the sea-sky line detection is more efficient and accurate. Finally, the Hough line detection and the least square method are used. Linear fitting is used to get the final sea-sky lines.

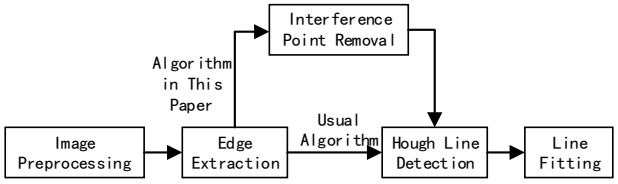


Figure 3. Overall flow chart

Step 1: We preprocess the image to solve the interference problems caused by strong watermark and uneven illumination. In which, the structural elements of closed operation and open operation are square structural elements.

Step 2:We use the Canny operator to extract the preprocessed sea- sky pictures.

Step 3: The mathematical morphology is carried out to remove the interference points, so that the sea-sky line detection is more efficient and accurate. The structural elements is the linear structural elements.

Step 4: We use Hough straight line detection and the least square fitting method to get the sea-sky lines.

The followings focuse on mathematical morphology of image preprocessing, interference point elimination, as well as the Hough line detection and other major steps to describe.

A. Image Preprocessing

In the process of the initial image collection, due to optical system distortion, relative motion, weather and other reasons, the noise is inevitable. In the process of transmission, noise can pollute the image and noise points have a certain bad effect on edge detection [11]. Generally, the Gauss filter is used to remove noise. However, this paper uses the mathematical morphology filter to preprocess the image, which can make the brightness of the target image more uniform and remove the interference of the watermark. At the same time, that can preserve the structure gradient information of the image better. In this paper, a closed-open filter(COF) is constructed through the combination of opening and closing. The definition is shown in the formula (5).

$$COF(f) = (f \bullet b) \circ b \tag{5}$$

The mathematical morphological filters have the properties of transitivity, translation invariance, idempotency and duality. The structure element chooses the larger square structure elements, because the noise element of uneven illumination and the strong watermark is larger.

B. Edge Detection

The edge detection by the Canny operator is a technology to extract useful structural information in different visual objects. And that greatly reduces the amount of data to be processed. It is now widely used in various computer vision systems. The Canny operators are used in different visual systems to detect edges, but the requirements on the edge detection are similar, so the wide application of edge detection technology can be realized[12].

For the grayscale image that has been preprocessed by the method of step 1, the gradient intensity and direction of each pixel in the image are calculated. The edges of the image can be directed to all directions, so the Canny algorithm uses four operators to detect the horizontal, vertical and diagonal border in the image. The operator of edge detection returns the first order value of horizontal Gx and vertical Gy direction, thereby the gradient G and direction theta of pixels can be determined, as shown in formula (6) and formula (7).

$$\mathrm{G} = \sqrt{\mathrm{G_x} + \mathrm{G_y}} \tag{6}$$

$$\theta = \arctan\left(\mathrm{G_x}/\mathrm{G_y}\right) \tag{7}$$

In formula (6),G is gradient strength. In formula (7), the arctan is an inverse tangent function.Besides, the theta represents gradient direction

The Non-Maximum Suppression is applied to eliminate the spurious response that is caused by edge detection. The Double-Threshold detection is used to determine the real and potential edges. Finally, the edge detection is completed by suppressing isolated weak edges.

The Canny operator is used to extract edges, so that all possible edges can be obtained to ensure the accuracy of edges.

C. Interference Point Removal

After using the Canny operator to extract the edges, the images with two value are obtained. There are still a lot of small noise points in the the images with two value. That causes interference to the next detection and fitting of sea-sky lines. Therefore, we use the mathematical morphology operation to remove interference points.

Because the mathematical morphological operation is sensitive to the size and shape of structural elements, the appropriate structural elements must be selected. Meanwhile, because the target is to obtain a sea-sky line, so the linear structure element is used to remove the noise points, so that the interference points can be removed .At the same time, the edge points of the target are not mistakenly removed [13]. The linear structural elements used in this paper are shown in the formula (8).

$$se = strel('line', x, y) \tag{8}$$

In formula (8), these character se represents the structure element, and the strel() is the function that created the structure element. In which, the 'line' represents a linear structure element, Meanwhile, the x and y determine the size and direction of the structure element that we choosed. Of which the linear structure element x and y have the following relationship, as shown in formula (9):

$$\begin{cases} x = 2N + 1 & N = 1, 2, 3, \ldots \\ q = 90 \ / \ (n - 1) & \theta \text{ is the unit Augle} \\ y = n * q & n = 0, 1, \ldots 4N - 1 \end{cases} \tag{9}$$

According to the characteristics of sea-sky line, this paper selects the linear structural elements, which can effectively remove interference points ,reduce amount of calculation in Hough line detection and improve its accuracy and efficiency.

D. Line Detection

The basic idea of the Hough transformation is the duality of the point to the line. After the image transformation, the images in the image space are transformed into the parameter space [14]. In the x-y image space, a straight line $y = Ax + B$ (Of which, A is the slope, B is intercept) corresponds the points in the ρ-θ parameter space.

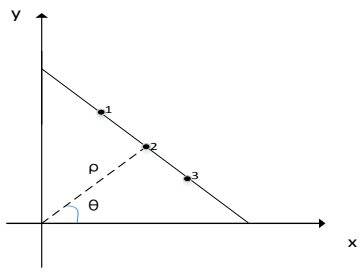


Figure 4. The image space

In the image space, the point on a straight line is a sinusoidal curve in the Hough parameter space; Many points on the same line in the image space are a sinusoidal cluster in the Hough parameter space and the curve clusters are intersected to a point, which is called the peak point. The peak point in the Hough parameter space corresponds to a straight line in the image space. As shown in Fig.4, this is the image space; As shown in Fig.5, this is the parameter space. The Hough transformation is converted from the image space of Fig. 4 to the parameter space of Fig.5.

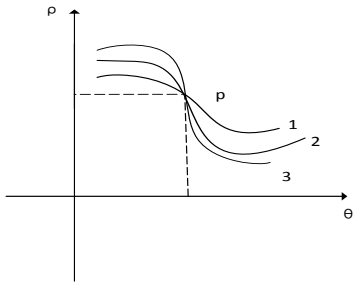


Figure 5. The parameter space

Therefore , Therefore, the Hough transformation transforms the straight line detection problems in the image space to the points detection in the parameter space. There are a number of possible lines in the sea-sky image, but the sea-sky line is throughout the image .Thus, the sea-sky line is The longest line segment in the sea-sky image, corresponding the local maximum value in the Hough parameter space. By detecting the local maximum in the Hough parameter space, we can find a corresponding line in the x-y image space, that is, the sea-sky line [15] .

E. Straight line fitting

Through the Hough Line detection, the longest line segment is extracted. However, the sea sky line is a straight line through the whole picture. So we have to extract and fit the points in the line segment and get the final sea-sky line. through the whole image. The sea-sky line is gotted by selecting some points and making straight line fitting by the least square method. In this paper, the least square method is used to fit the straight line. The least square method is a mathematical optimization technique. It searches for the best function matching of data by minimizing the sum of squares of errors. The least squares method can be used to obtain the

unknown data simply and make the sum of squares between the obtained data and the actual data minimum.

IV. Experiment and Result Analysis

To verify the results of this method, we selected three sea-sky background images in different environments, as shown in the Fig.6 .The Fig. (a) is a sea-sky image with lower visibility; The Fig. (b) is a sea-sky image with strong

watermark; The Fig. (c) is a sea-sky image with uneven illumination. After the operations are performed on the matlab 2015a software, two mathematical morphological processing cases are compared and analyzed.

The mathematical morphology operations are used in the pretreatment of sea-sky image.Two preprocessing methods are used. One way is to use the Gauss filter to reduce noise and then conduct sea-sky lines detection. The results are shown in Fig.7 . The other way is using the mathematical morphological filter to reduce noise and then conduct the sea- sky line detection, the results are shown in Fig.8 . As seen from the Fig. 7(b), the former method is not ideal for detecting sea-sky pictures with strong water marks, and there is an error. From the fig. 8 , we can see that the method of this paper can accurately detect the sea-sky line in different environment.

(a) (b) (c)

Figure 6. Original picture of sea-sky background

(a) (b) (c)

Figure 7. Sea-sky-line detected after Gauss filter processing

(a) (b) (c)

Figure 8. Sea-sky-line detected after mathematical morphological processing

The images are results of sea-sky lines detection in Fig.8.The sea-sky-line detected by the algorithm in this paper .In order to measure the processed image quality, we usually refer to the PSNR value to determine whether a particular processing program is satisfactory enough.Thus, to quantitatively evaluate the experiment results, the experiment performance of the sea-sky line detection in three sea-sky backgound images are compared. The peak signal to noise ratio (PSNR) is used. The PSNR is the ratio of the variance to the information and noise, when the value of

PSNR is between the 30dB and 40dB,which means less noise. When the value of PSNR is 40dB ,which means better picture processing effect [16]. The expression of the PSNR is shown in the formula(10).

$$PSNR = 10 \times \lg \frac{f_{\max}{}^2}{MSE} \tag{10}$$

In which, $f_{\max}$ is the maximum grayscale value of function $f(x,y)$. MSE is a mean-square error that reflects the variance between the estimate and the estimated amount, as shown in the formula (11).

$$MSE = \frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[g(x,y) - f(x,y)]^2}{MN} \tag{11}$$

The obtained quantitative evaluation experiment data are shown in Table 1.

TABLE I. COMPARISON OF PNSR VALUES OF EXPERIMENTAL RESULTS

	Image processed by Gaussian filter		Image processed by filters in This paper	
	MSE	**PNSR /dB**	**MSE**	**PNSR /dB**
Fig.6 (a)	5.5771	37.6667	9.2155	38.4856
Fig.6 (b)	17.3281	35.7433	26.4623	33.9045
Fig.6 (c)	74.8766	29.3873	36.0083	32.5571

From the numerical change of PSNR in Table 1, we can see that the quality of the picture is in high level in most cases ,for the sea-sky background pictures are processed by the Gauss filter. However, from Fig. 6 (c), we can see that the value of PNSR is lower than 30 dB. Thus, for the unevenly illuminated sea- sky background pictures, the Gauss filter can not effectively process and solve the uneven illumination problem; For Fig. 7 (b), the PSNR values of the pictures after the Gauss filter processing are higher than the result of the pictures after the mathematical morphological filter processing. As a result, it can be seen that the quality the pictures after the Gauss filter processing is relatively higher than that after morphological filter processing. It is also shown that the Gauss filter can effectively remove the interference of strong watermarks on the image processing. However, according to the Fig.6 (b) effect diagram of the sea-sky lines detection after the Gauss filter processing, it is known that although the pictures are processed by the Gauss filter. They get better quality of the images. It effectively removes the interference of the strong watermarks, but also loses more detail information of the edge of the sea and the sky, so that the effective points of the edge are also erroneously removed. There is not an accurate sea-sky line to be obtained . As seen from the Table 1, the PSNR values of the images that are processed by the mathematical morphological filter in this paper are all between 30 dB and 40 dB, which indicates that the image processing quality is better. At the same time, combining with figure 8, we can know that the algorithm in this paper can effectively preserve the details of the sea-sky boundary by the algorithm in this paper. The more accurate sea-sky lines can be obtained. The sea-sky lines in different sky-sea background images are measured. It objectively reflects the feasibility and superiority of the algorithm in this paper.

The mathematical morphology is applied to the interference point removal after edge detection by the Canny operator .After the interference points are removed, the Hough transformation is used to extract the sea-sky lines so as to achieve the final detection and fitting of the sea-sky lines .We get the detection result of the sea-sky lines finally,as shown in Fig. 8 . We obtain the two-value pictures after edge detection by the Canny operators.Thus, the mathematical morphology is applied to the two-value pictures.The experimental data of Table 2 are analyzed by comparing the number of effective points reserved before and after second mathematical morphology processing in the experiment.

TABLE II. NUMBER COMPARISON OF VALID POINT BEFORE AND MATHEMATICAL MORPHOLOGY PROCESSING

	Possible points A	**valid points B**	**Ratio b/a**	**Time-consuming /ms**
Fig.6 (a)	100	28	28%	688
Fig.6 (b)	590	252	42.7%	674
Fig.6 (c)	102	72	70.6%	708

From table 2 , it can be seen that after the second morphological processing ,based on the special structure element, the interference points of the sea-sky lines can be reduced effectively. Because of the edge extraction using the Canny operator, all the possible edges are detected. However, only one of the required sea-sky line is needed. At the same time, some interference points are produced when the edge detection is performed. These are unavoidable. Under the premise of guaranteeing the valid the sea-sky line information point ratio (B/A), the effective detail of sea-sky lines are protected, the possible points (A) are effectively reduced . From the data change in Table 2, we can see that the interference information produced by edge extraction is the lower for images with lower visibility. For images with uneven illumination, there are more interference information produced by the operation of preprocessing and edge extraction. These interference information will cause the interference problems on straight line detection and fitting. That makes the sea –sky line extraction difficult and inaccurate. At the same time, it shortens the time of straight line detection and improves the efficiency of the algorithm in this paper. The mathematical morphologic is used, the computation amount of Hough detection is reduced. Meanwhile, the algorithm in this paper reduces the interference on the sea-sky line fitting. Thus, the algorithm in this paper ensures the efficiency and accuracy of the sea-sky line detection.

Therefore, the method of this paper achieves the expected effect and the extraction effect of sea-sky lines is ideal.

V. CONCLUSION

This paper presents a method of the sea-sky line detection based on the mathematical morphological. Firstly, the image is preprocessed by mathematical morphological filtering. Followed that, the Canny operator is used to extract the sea-sky boundary. Secondly, the Mathematical morphology processing is once more used to remove the interference points on the sea-sky line; Finally, the sea-sky line is detected by the Hough transform and fitted by the least square method. The experimental results show that, this algorithm can detect the sea-sky lines, as well as the robustness is better, accuracy is higher. It can effectively cope with the complex marine environment and weather effects.

REFERENCES

[1] Messages , Zhengjia , under combined .Sea-sky line detection algorithm based on morphological processing and least square method [J]. Optics and Optoelectronic Technology , 2013, one (1): 000091-94.

[2] Liang D, Zhang W, Huang Q, et al. Robust sea-sky-line detection for complex sea background[C]// IEEE International Conference on Progress in Informatics and Computing. IEEE, 2016:317-321..

[3] H. Wang, Z. Wei, S. Wang, et al. A Vision － based Obstacle Detection System for Unmanned Surface Vehicle［C］/ /Proceeding of the 2011 IEEE Conference on R obotics , Automation and Mechatronics, 2011: 364 ～ 369.

[4] Wang Bo, Su Yumin, Wan Lei, et al. Sea sky detection method based on gradient saliency for surface unmanned craft [J]. Acta optica Sinica, 2016 (5): 66-75.

[5] Dryden , Zhang . Xinggang. A method of periodic noise removal based on weights adaptive Morphology [J/ol]. Computer technology and Development , 2018 (a): 1-8[2018-04-25].

[6] Jiang L, Guo Y. Image Edge detection based on adaptive weighted morphology[j]. Chinese Optics Letters, 2007, 5 (2): 77-78.

[7] Feng Gui, GUI pre- , , Wind, Lin . . Morphological method in edge detection of gray image [J]. Remote sensing information , 3:12-14,.

[8] Wanghuifeng , War Guile , Luo . Xiaoming. Research and application of edge detection algorithm based on mathematical morphology [J]. Computer Engineering and Applications , 2009 (9): 223-226.

[9] Zhong Junliang. . Real-time detection and tracking technology of infrared small and dim targets [D]. Graduate School of Chinese Academy of Sciences (Changchun Institute of Optics and Fine mechanics and Physics , 2013.

[10] Wang Fang , Changwei , Li . Wenshu. Image edge extraction method based on mathematical morphology [J]. Mechanical Engineering and Automation , 2015 (1): 46-48.

[11] Li Mu, Chi ge, etc. Adaptive Canny operator Edge detection technology [J]. Journal of Harbin Engineering University, 2007, (9): 1002-1007.

[12] Wang Guangling. . Research on the algorithm of detecting and tracking video velocity based on moving objects [D]. Taiyuan University of Technology , 2009.

[13] Zhanghuang , , Yu Shenglin, Bai Bangong. selecting principles for structural elements in morphological image denoising [J]. Data acquisition and processing , 2008, S1:81-83.

[14] Dong Yu Star, Liu , Weining, dongyu-xing, , . . Small target detection for Sea-sky background based on Gray character [J]. China Optics , 3 (3): 252-256.

[15] Lu Junwei, Wang , Xiaodong,, and . on. Sea-sky line detection algorithm based on fractal feature and Hough transform [J]. Journal of the Naval Institute of Aeronautical Engineering , 2006 (5): 545-548.

[16] Xu Tianji , Zhang Guican, Jack. , . Watershed color Image segmentation based on morphological gradients [J]. Computer Engineering and Applications , 2016, (one): 200-203

Cheap and Simple Slip Sensor Based on the Optical Sensor

Feibi Liu
University of Southampton, Southampton, UK, SO17 1BJ
e-mail: 2496909486@qq.com

***Abstract*—The purpose of this project is to manufacture a cheap and simple slip sensor. The slip sensor is based on the optical sensor. The experiment divides into two parts, how to build sensor and how to test the features of the sensor. The slip sensor could be divided into three parts, cover, optical sensor and adaptor board, which will connect to the suitable resistors circuit to obtain the large gain. During the test, the participant applies the static state way to measure the relationship between input and output with a white cover and a gray cover. The slip sensor with the white cover has low hysteresis and high repeatability which are better than the gray cover.**

Keywords-Slip Sensor; Optical Sensor; Force Sensor; Static State; Reflective Object Sensor

I. INTRODUCTION

When humans try to grasp an object, they do not need to know the parameters (like mass) of the object. They just control the force of the hand which ensures the object will not fall by human 'feeling'. For the robot hand, the slip sensors which fix on the fingertips could give the feedback (it is like human's feeling) to the CPU which could control the force of the robot hand [1]. In a sense, the slip sensor is a kind of the force sensor, because the feedback of the slip sensor is related to the force which puts on the surface of slip sensor. Only different is the slip sensor focuses on the force changing between the object slipping and not slipping.

There are many kinds of methods to build a slip sensor or a force sensor, which bases different theories. At the beginning, Luo used piezoresistive strain gauge on the robot links to detect the force and this was used in this field for a long time. However, this structure was a little complex comparing the new ways [2]. Cristina Cristalli and Michael R Neuman designed a kind of force sensor by changing the capacitance in dielectric structure[3].They used this sensor to measure the blood pressure and gain a good ratio between the input and output. However, in the experiment, they also found that the ratio between the input and output would change a little when they used the same way to build a same standard sensor. The reason is the different stray capacitances in the environment. Piezoresistive is very popular in industry and research field because of low cost, small and light-weight. However, piezoresistive exists low repeatability and large hysteresis, which will cause the low accuracy. L.Paredes-Madrid group improved the accuracy of piezoresistive by modeling the capacitance. Nevertheless, they still needed to consider how to reduce the force estimation errors [4]. Lorenzo Jamone and his group designed a kind of tactile sensor that based on the hall-effect theory. This sensor detected the changing of the magnetic field when the force pushed on the sensor to give a feedback to control. After the experiment, they found this kind of sensor had high sensitivity, low hysteresis, and good repeatability, but the experiment just tested the force on normal component [5].

Samir Boukhenous and Mokhtar Attari also used the Hall-effect theory to produce a pinch grip sensor, which had a good response to input and output. However, the experiment also focused on the normal direction [6]. Darko Belavic and his group completed an experiment about the low energy consumption of different types of pressure sensors which included the piezoelectric sensor. This piezoelectric sensor was based on the performances of the ferroelectric thick films, which could transform the pressure into shifted resonant frequency of diaphragm which could be regarded as an output signal. They concluded that the piezoelectric resonant sensor was suitable for low energy consumption and the energy consumption was mainly

dependent on materials and structures, but it was not sample work to reduce consumption because of complete relationships between every component in sensor [7]. Shouhei Shirafuji and Koh Hosoda developed a robot hand which combined the strain gauge sensor and piezoelectric polyvingylidenefluoride (PVDF) sensor to measure stresses and detect slipping. The PVDF sensor could detect the variations of the force which was caused by the slipping of the object. However, the PVDF sensor needed the high reactive ability and high resolution to detect the slipping [8]. A. Persichetti, F. Vecchi and M. C. Carrozza, presented a contact sensor using optical theory, which based on detecting the changes of the light beams intensity. Figure 1 showed the structure of this kind of sensor, which included a soft silicone cover, a receiver (a phototransistor) and a transmitter (an infrared photodiode). When the force pressured on the cover, the intensity of the light which reflects the receiver will change. Therefore, this sensor has high sensitivity, fast response and could enhance the immunity to noise by adjusting the light intensity from the transmitter [9].

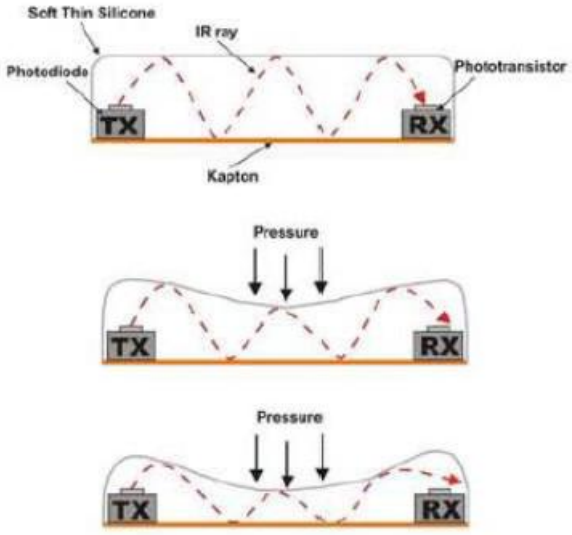


Figure 1. The structure of the optical sensor

For this paper, the participant will utilize the optical theory to build a cheap, simple and low consumption slip sensor. Because the photodiode needs to emit high intensity light, the optical sensor will cost high consumption compared with other sensors [10]. Therefore, the participant tried to find a good material which could reduce the loss during reflecting, which could reduce the requirement of the light intensity which is emitted by the photodiode. This could reduce the consumption. For obtaining a good response, the participant also chooses suitable resistors to get large amplifier gain. Comparing with an optical sensor which is built by A. Persichetti and his friends [9], the most different in this paper is the participant use cast Perspex acrylic sheet to replace silicone (showing figure 1) as the cover. Because the cast Perspex acrylic sheet is harder than silicone, this replacement could increase the robustness of the sensor. However, this also means the slip sensor may sacrifice some sensitivity.

II. METHODOLOGY

The purpose of the experiment is to produce a cheap, simple and low consumption slip sensor. Therefore, the experiment could be divided into two parts, the first part is to build the slip sensor and the second part is to test the slip sensor as figure 2.

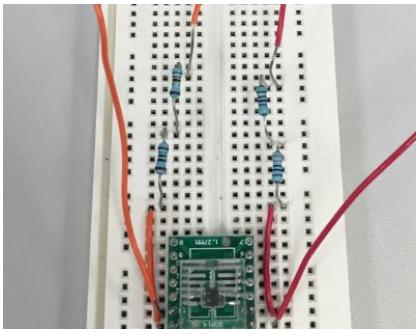


Figure 2. The slip sensor with resistors

A. Design

The slip sensor includes three parts, cover, optical sensor and adapter board. Figure 3 shows the slip sensor. Therefore, first step, the participant needs to choose a suitable optical sensor.

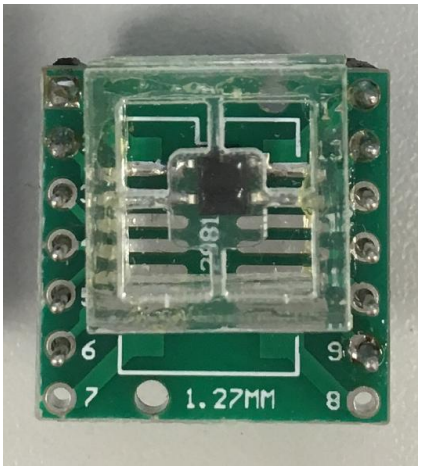


Figure 3. The slip sensor

The QRE1113RG-Minature Reflective Object Sensor is used as the optical sensor because it is mini. As Figure 4 showing, the photodiode (transmitter) which is between the pin 1 and pin 2 emits the light and the phototransistor (receiver) which will receive the reflected light from the cover is fixed between pin 3 and pin 4 [11].

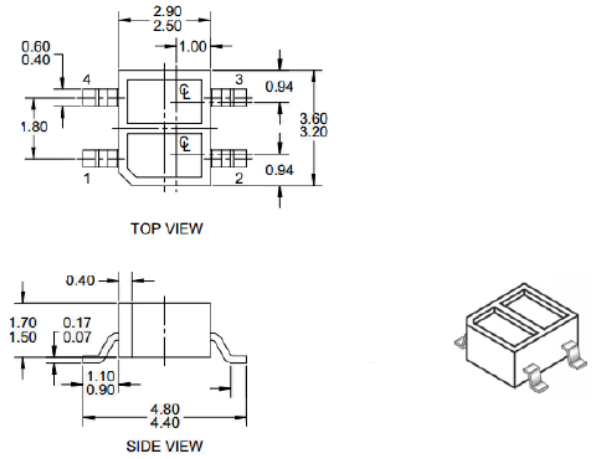


Figure 4. The structure of the QRE1113RG sensor

Second step, the participant needs to produce a suitable cover. The cover is made by the 1mm thickness clear cast Perspex acrylic sheet. There are some advantages to using this material. Firstly, the weight is light. Secondly, the cost of the material is low. Thirdly, it is easy to build which just needs a laser cutter. Fourthly, the cast Perspex acrylic sheet has high tensile strength and rigidity, so it could protect the optical sensor which is under the cover. Fifthly, the clear Perspex acrylic could transmit 92% of the visible light, which is simple to the participant to change the color of the cover and detect the influence of the color of the cover [12]. Figure 5 shows the blueprint of the cover. Figure 6 shows the cover after gluing.

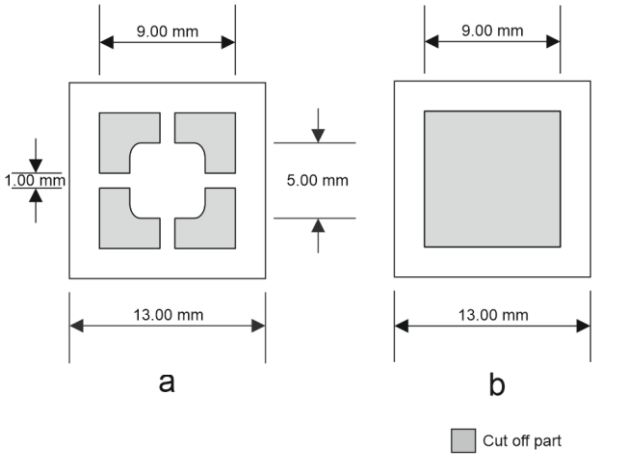


Figure a and b are axial symmetry and central symmetry

Figure 5. The cover's blueprint

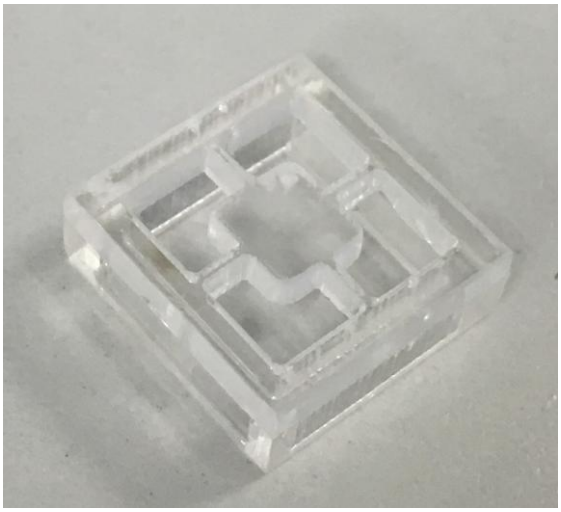


Figure 6. Cover

Third step, solder the optical sensor on the adapter board as the figure 7.

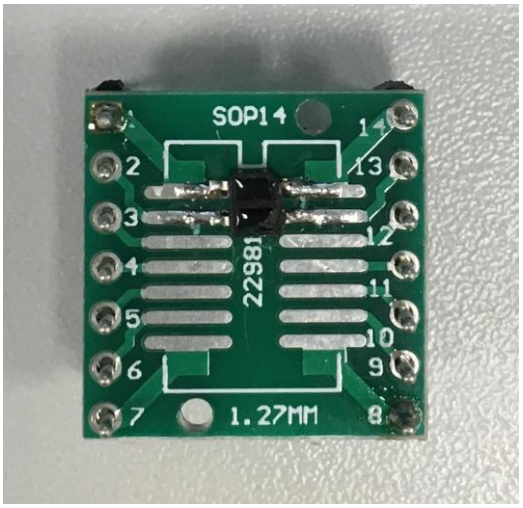


Figure 7. The optical sensor with adaptor

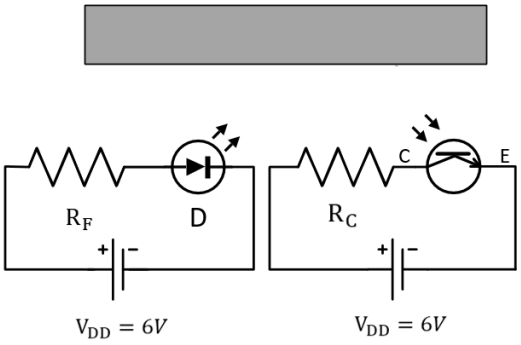


Figure 8. The circuit's schematic structure

Fourth step, choose the suitable resistors to build the circuit. Figure 8 shows the equivalent circuit diagram. RF is used to keep the photodiode (transmitter) working in the ideal voltage range. RC is to amplify the voltage signal. According to the specification of the QRE1113RG, the participant assumes IF =0.028A, which could ensure the photodiode work well at room temperature.

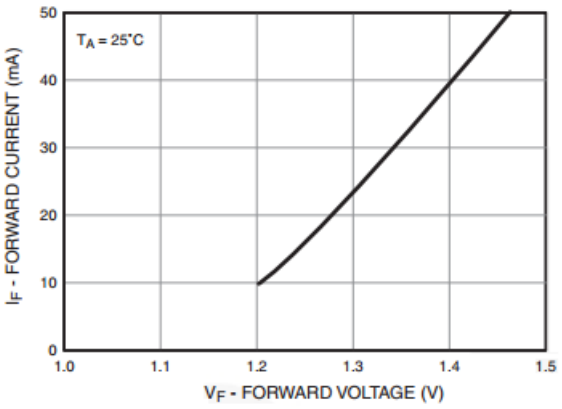


Figure 9. Forward current VS Forward Voltage

From figure 9, the forward voltage (VF) should be about 1.33V. According equation 1,RF=168Ω.

$$V_{DD} = R_F I_F + V_F \quad (1)$$

Because the participant wants to get the largest reflection light from the cover, the participant puts a white thin paper on the top of the cover during the test. The distance between the top surface of the cover and the top surface of the optical sensor is about 2mm. According to the figure 10, it could easy find that the real collector current IC is 0.45 which is half of collector current at distance 1mm.

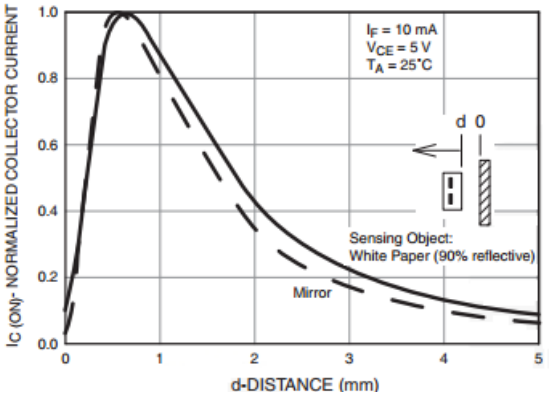


Figure 10. Normalized Collector Current VS Distance between

When IF =0.028A and distance is 1mm, the collector current is around 1.08mA in according specification, so the real collector current IC is0.54mA. Assume the collector-emitter voltage (VCE) is 3.3V which could get

larger gain. Using the equation 2 gets the collector resistor (RC= 5000Ω).

$$V_{DD} = R_C I_C + V_{CE} \quad (2)$$

B. Test

These devices were applied in the experiment. EL 155R Power Supply (6V and 0.028A);Amprobe 30XR-A multimeter; Compression testing machine;MDO4054B-3 Mixed Domain Oscilloscope;HMC 8012 digital Multimeter. Figure 11 shows the schematic during testing.

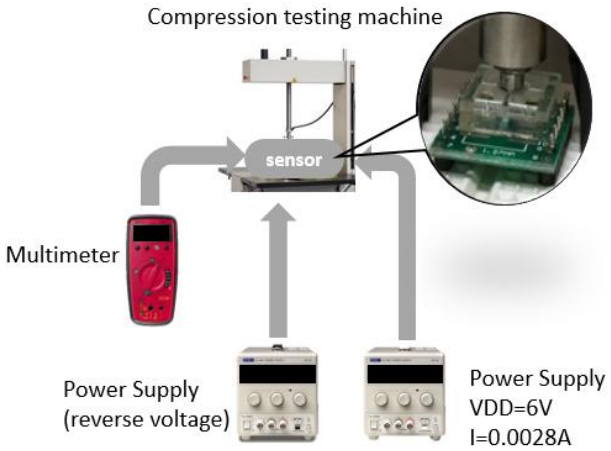


Figure 11. The assembly schematic

First step, the participant put a white paper on the cover to detect the relationship between the force and voltage. Because the output voltage was more than 2V which is exceed the range of the Amprobe 30XR-A multimeter, the second power supply gave 1V reverse voltage as the offset voltage. Second step, the participant put a weight (1 N) on Compression testing machine and recorded the voltage. Third step, repeated step 2 until there were six weights (6N) on the machine. Fourth step, took the weight (1N) one by one from the machine and recorded the voltage after each operation. After these steps, one group of data was completed. To obtain stable data, the participant repeated five times of these steps and got five groups of data.

Next, the participant changed the color of the cover into gray, which just put nothing between the probe and cover, because the probe is in gray. Then, used the same way to get five groups of data. The output voltage is not more than 2V, so the offset voltage does not need. The second power supply is 0V.

III. RESULT AND ANALYSES

Figure 12 and figure 13, show the relationship between the output voltage and force using white cover. The trend line in each figure could quantize the relationship between the output and input. The R-square shows on each figure mean the reliability of the trend line. When the value is more near the 1, the points should more near the trend line. Usually, when the value is larger than 0.75, the trend line could predict the relationship between the input and output. In each of figure 12 and figure 13, the R-square is larger than 0.98, which means the trend line has high predictability. From these two figures, it is easy to find that the coefficients of these two trend lines are close. Therefore, the sensor has low hysteresis and high repeatability in white color. The offset voltage is around 2.77V, which is a little high. The reason is that the white color reflects almost light which enhances the power that phototransistor received.

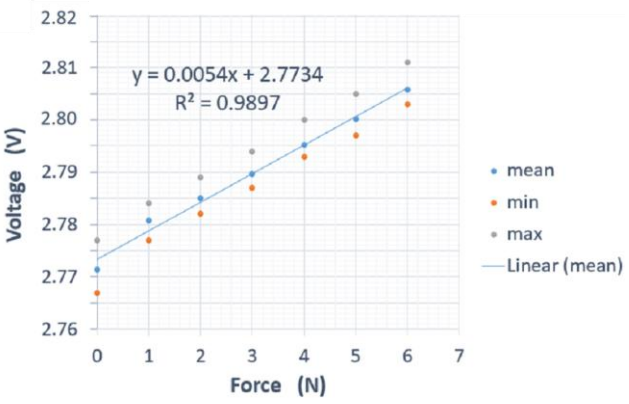


Figure 12. Meaning of Voltage Vs Force when force is increasing

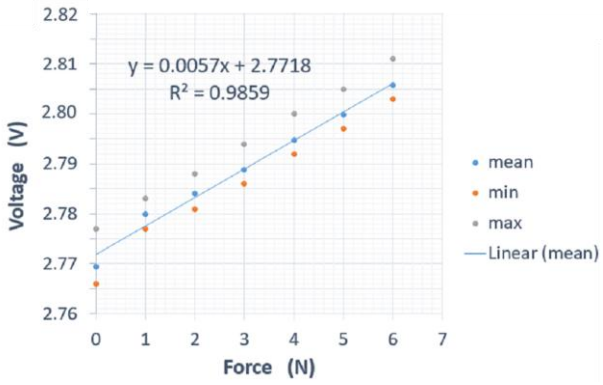


Figure 13. Meaning of Voltage Vs Force when force is reducing

Figure 14 and figure 15 show the relationship between the output voltage and force using gray cover. However, it is easy to find that the R-square in these two figures is much lower than figure 12 and figure 13. The reason is that the features of the optical sensor between collector current and distance getting from the specification are based on the white paper. All assumes are also based on that condition, using white paper (90% reflective). Therefore, the accuracy of the trend line is low during using gray cover. However, the gain using gray cover (around 0.01) is two times than the gain using white cover (around 0.005). The participant speculates one of group data in figure 14 and 15 is not reliable. From the figure 14 and figure 15, the max output points are much higher than average output points, because one group of the value is much higher than others. This may be caused by hysteresis. Nevertheless, to ensure the authenticity of the experiment. The participant still recorded this group of value.

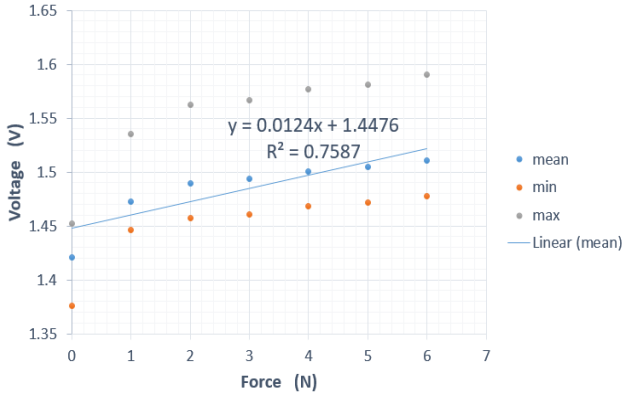


Figure 14. Meaning of Voltage Vs Force when force is increasing

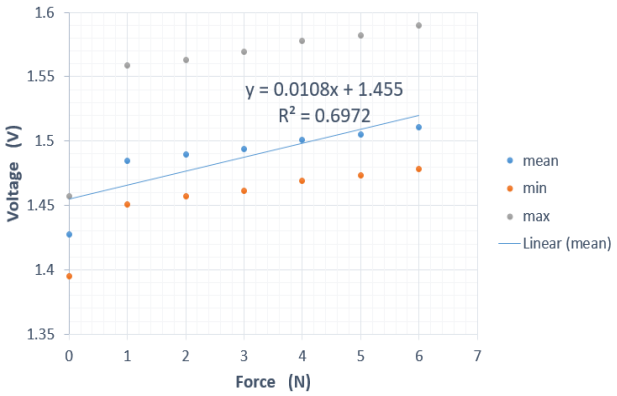


Figure 15. Meaning of Voltage Vs Force when force is reducing

IV. CONCLUSIONS

During the experiment, the participant used a cheap and simple way to build a slip sensor. The participant also found that white cover of the sensor has high repeatability, low hysteresis, and high predictability. However, the gray cover of the sensor has low repeatability and low accuracy. Comparing with other slip sensors which use the optical theory, this slip sensor has high robustness.

In the test, the participant used static state way to test the feature of the slip sensor, which could obtain the quantized function information between input and output. However, in the future work, the participant could improve the experiment from these two aspects. First one, the participant should try to detect the response of slip sensor in the dynamic state which means the object is slipping on the sensor. Second one, the participant should design a black box to control the external light. This could help the participant understand the influence of the light in the environment to the sensor cover and find a high resistance of cover to the external light.

REFERENCES

[1] Paul H Chappell, Darryl P J Cotton, Andy Cranny and Neil M White "EXPERIMENTAL LEAD ZIRCONATE TITANATE (PZT) SLIP SENSOR", MEC '08 Measuring Success in Upper Limb Prosthetics, Aug.2008

[2] R. Luo, "A Microcomputer-Based intelligent sensor for Multiaxis Force/Torque Measurement", in IEEE Transactions on Industrial Electronics, vol. 35 pp. 26-30. February 1998.

[3] C. Cristalli and M. R. Neuman, "A capacitive pressure sensor for measurement of interfacial pressure between a sphygmomanometer cuff and the arm," Engineering in Medicine and Biology Society,

1995., IEEE17th Annual Conference, Montreal, Que., 1995, pp. 1541-1542 vol.2.

[4] L. Paredes-Madrid, L. Emmi and P. Gonzalez de Santos, "Improving the performance of piezoresistive force sensors by modeling sensor capacitance," 2010 IEEE International Symposium on Industrial Electronics, Bari, 2010, pp. 458-463.doi: 10.1109/ISIE.2010.5637681

[5] L. Jamone, L. Natale, G. Metta and G. Sandini, "Highly Sensitive Soft Tactile Sensors for an Anthropomorphic Robotic Hand," in IEEE Sensors Journal, vol. 15, no. 8, pp. 4226-4233, Aug. 2015. doi: 10.1109/JSEN.2015.2417759

[6] S. Boukhenous and M. Attari, "An easy made pinch grip sensor to quantify fingertip pressure," 2008 2nd International Conference on Signals, Circuits and Systems, Monastir, 2008, pp. 1-3. doi: 10.1109/ICSCS.2008.4746929

[7] D. Belavic et al., "Low energy consumption thick-_lm pressure sensors," Microelectronics and Packaging Conference, 2009. EMPC 2009.European, Rimini, 2009, pp. 1-6.

[8] S. Shirafuji and K. Hosoda, "Detection and prevention of slip using sensors with different properties embedded in elastic artificial skin on the basis of previous experience," Advanced Robotics (ICAR), 201115th International Conference on, Tallinn, 2011, pp. 459-464.

[9] A. Persichetti, F. Vecchi and M. C. Carrozza, "Optoelectronic Based Flexible Contact Sensor for Prosthetic Hand Application,"2007 IEEE 10th International Conference on Rehabilitation Robotics, Noordwijk, 2007, pp. 415-420.

[10] P. H. Chappell, "Making sense of artificial hands," Journal of Medical Engineering Technology, 2011, vol.35

[11] "QRE1113 PDF Datasheet Reflective Object Sensor", Fairchildsemi.com, 2002. <https://www.fairchildsemi.com/products/optoelectronics/infrared/reflective-sensors/QRE1113.html?keyword=QRE1113GR.> (01-Sep-2016).

[12] "Perspex Design Guide - guidance for fabricators and designers. - Perspex", Perspex.co.uk. <http://www.perspex.co.uk/technical-library/brochures-and-workbooks/perspex-design-guide/> (01- Sep- 2016).

Research and Improvement of Apriori Algorithm Based on Hadoop

Gao Pengfei [a], Wang Jianguo [b] and Liu Pengcheng [c]
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, Shaanxi Province, China
[a]gaopf1225@gmail.com, [b]wjg_xit@126.com, [c]294843945@qq.com

***Abstract*—Association rules can forcefully get a horizontal relation in the big data, the Apriori algorithm is one of the most significant association rules. Traditional mining based on parallel Apriori algorithms needs much more time in data IO with the increasing size of large transaction database. This paper improves the Apriori algorithm from compressing transactions, reducing the number of scans and simplifying candidate set generation. And then the improved algorithm is parallelized on the Hadoop framework. The experiments show that this improved algorithm is suitable for large-scale data mining and has good scalability and effectiveness.**

Keywords-Apriori algorithm; Hadoop; Association rules

I. INTRODUCTION

In the context of the development of big data "spraying wells", there is frequently a close relationship between vast amounts of data[1]. Analysis and decision making through data mining have become the mainstream of social development. In order to better find the relevance of transaction data sets, some researchers have discovered the concept of association rule mining technology[2]. With the attention of many researchers at home and abroad caused by the conception of the concept, they have done a lot of analysis in this field and put forward many data mining algorithms.

One of the most famous association rule algorithms is the Apriori algorithm, which is a classic association rule algorithm designed by Agrawal[3-4] in 1994. It is a level-by-level search iteration method that constructs a k-item set to constitute a k+1-item set. The main ideas of this algorithm are: Firstly, all frequency sets are counted from the transaction database, and the support of this frequent set must not be less than the minimum support degree; Secondly it enters into the process of strong association rule generation, and the rules need to satisfy the support and confidence thresholds at the same time; Thirdly, only all rules that contain collection items are retained. Once these rules are retained and generated, that are greater than or equal to the MinConfidence.

The design of the Hadoop[5] framework originated was from an open source project developed by the Apache organization Foundation. Because of its inter-temporal significance, the Hadoop framework has been widely used in the information field at home and abroad. There are two important modules in the Hadoop frame--Distributed File System HDFS and Distributed Computing Frame MapReduce[6]. As a distributed file system , HDFS functions aims to implement data storage. It will work in conjunction with the computational framework. MapReduce works to provide the underlying support for data calculations; And the idea of MapReduce[6-7] is based on a paper by Google. In short, its core method is "the decomposition of tasks and the statute of results."

II. BRIEF AND RESEARCH STATUS OF APRIORI ALGORITHM

A. Overview of Apriori algorithm

The Apriori algorithm is a level-by-level search iterative method that consists of a k-item set to construct a (k+1)-item set. First, obtain a frequent 1-item set. L1 can generate a frequent 2-item set L2, and L2 can generate a frequent 3-item set L3. According to this rule, when a frequent k-item

set cannot be found, the algorithm ends[8-9]. The specific operation is as follows:

1) Iterate through the initial transaction database and count the frequency of occurrence of the candidate set. The result is the support of the project. All projects whose all supports level no lower than the preset threshold generate a frequent 1-item set L1.

2) The algorithm uses L1 JOIN L1 to form a candidate C2-item set C2.

3) Using the items in C2, traverse the database again to obtain the support degree of each candidate set. All projects with support levels not lower than the support level generate frequent 2-item set L2.

4) The algorithm uses L2 JOIN L2 to form a set C3 of candidate 3-item sets.

5) Using the items in C3 to traverse the database again, the support degree of each candidate set can be obtained. All items with support levels not lower than the support level generate frequent 3-item set L3.

The above process is performed iteratively until the candidate set C k is empty. The Apriori algorithm does multiple IO operations on the database. Each stage consists of two parts, namely connection and pruning.

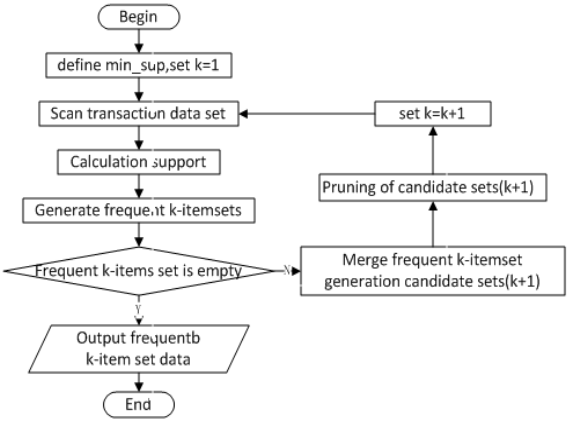


Figure 1. Apriori flow

B. Apriori algorithm instance analysis

Original transaction T10={A,C,D}, T20={B,C,E}, T30={A,B,C,E}, T40={B,E}. Suppose min_sup=2. Then L1={{A},{B},{C},{E}}, L2={{A,C},{B,C},{B,E},{C, E}}.

1) Self join, C3=L2×L2={{A,C},{B,C},{B,E}, {C,E}} × {{A,C},{B,C},{B,E},{C,E}}={{A,B,C},{A,C,E},{B,C,E}}

2) Pruning, Any frequent item set, its subset must also be frequent. For Candidate Set C3, clearing those subsets with infrequent options: The two items subset of {A,B,C} are {A,B},{A,C},{B,C}, where {A,B} is not an element of L2, so remove this option.; The two items subset of {A,C,E} are {A,C},{A,E},{C,E}, where {A,E} is not an element of L2, so remove this option; All the two items subset generated by {B,C,E} are {B,C},{B,E},{C,E}, the subsets produced by {B,C,E} all satisfy the requirements of L2. Therefore, this option is not deleted.

3) In this way, C3={{B,C,E}} obtained after pruning.

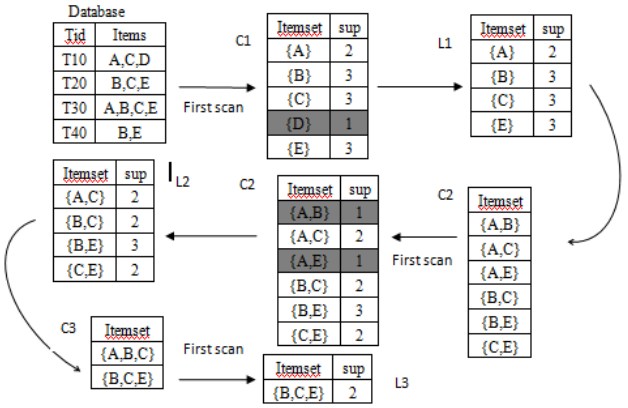


Figure 2. Apriori algorithm execution process

C. The shortcomings of Apriori algorithm

1) When the Apriori algorithm generates the candidate item set, it needs to perform the self-connection operation on the frequent itemsets obtained in the previous step. Then scan the transaction data set again and compare the candidate set formed by the self-connection with min_sup. During the self-connection operation, a large amount of comparison work will be performed.

2) Apriori algorithm need to rescan transaction datasets before pruning, and then compare with min_sup. Therefore, when the size of the transaction dataset is getting larger and larger, each scan will consume a lot of time, resulting in inefficient mining.

3) In the current situation where the data information has a high dimension and the type is complex, the classical Apriori algorithm can't satisfy users.

4) Because the classic Apriori algorithm is only applicable to a single machine, when the size of transaction data sets gradually becomes larger and larger, it will lead to inefficient mining, insufficient storage space, and even system crashes.

III. Parallel MEC-Apriori Algorithm Based on MapReduce Apriori algorithm

A. *Reduce frequent item sets self-connection comparison times and pruning steps*

In the processing of candidate sets, a method of transaction compression characteristics has been introduced. That is, according to the n-dimensional data item set, if itself is not a frequent item set, then the n-1 dimensional subset of the n-dimensional data item set is also not a frequent item set. Therefore, in the mining of candidate sets in the transaction database, the number of candidate candidate sets is compared and deleted because of the method of transaction compression characteristics, so that the number of candidate sets is gradually reduced, and the time efficiency of mining frequent itemsets is improved.

B. *Reduce the Number of Scanned Databases*

When mining frequent itemsets, the original transaction database is converted into a vertical data table, and then scan the vertical data table to mine frequent itemsets, because only one transaction database was scanned, a problem with frequent I/O was solved to some extent.

C. *Combining Apriori Algorithm and Hadoop Platform*

With the ever-increasing size of data, the traditional Apriori algorithm has been difficult to support its massive database of transactions. The solution to this problem is to add the Hadoop distributed platform to the Apriori algorithm[10], which not only makes the traditional Apriori algorithm run more efficiently, but also eases the storage pressure of the transaction database.

1) Generate frequent itemsets

The flow chart in this stage is shown in Figure.2

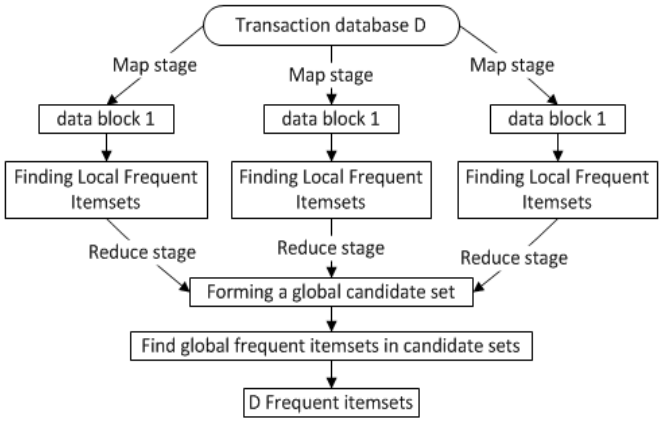


Figure 3. Generate frequent itemsets flow

a) The way of data blocks formatting

For the function of interface named Input Format implements Record Reader(Interface) is to convert data blocks into key-value pairs, eg: <key1,value1>.

b) Perform Map task

The idea of the first step is to generate the frequent item sets of each block.

c) Perform Reduce tasks

The key-value data output by the Combiner function is used as the input data of the Reduce phase. After a series of merging processes, some frequent item sets of the data module are obtained as a global candidate item set.

d) Scan *transaction* data set D

Call the Map function to rescan the formed global candidate frequent item set, and self- join , compare the minimum support count with the set of transaction items formed by the self-join, If it is less than the minimum support, then the last local frequent itemset is the final global frequent itemset, then pass it to the Reduce function and summarize it. Instead, it is necessary to iterate the local frequent itemsets until a frequent itemset is generated.

2) Generation of association rules

After association rules mine frequent item sets, it is necessary to generate strong rules. The emergence of strong rules is shown in Figure.3:

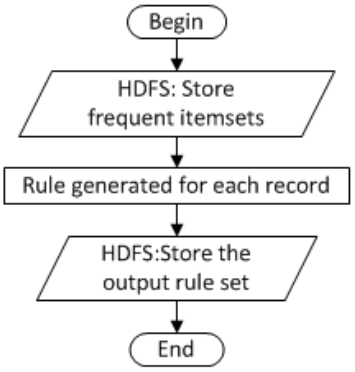


Figure 4. Generate strong rules flow

a) In the transaction dataset that holds the text, the input data of the Map must exist in the form of a key-value pair, so each row of data can be treated as a transaction. The key in a key-value pair is the offset of each row of data, and the value is represented as this row of data.

b) These key-value pairs are used as the input of the Map function, and then a set of frequent items conforming to the actual situation is obtained according to the set support threshold.

c) The output of Combiner function in Map stage is used as the input data of Reduce stage, then it is processed according to the local frequent itemsets generated in Map stage, and finally the strong association rules of the output are stored in HDFS.

IV. Experimental Assessment and Analysis

A. Setting Up a Hadoop Cluster Environment

The size of a Hadoop cluster is arbitrary. A small cluster can consist of a NameNode and several DataNodes. And a large cluster can consist of a NameNode and hundreds of DataNodes. Local mode, pseudo-distribution mode and fully-distributed mode are three modes built by Hadoop clusters. Considering the hardware configuration problem, This paper choses to use a virtual machine to set up a cluster environment, and the number of nodes in the cluster is 3, as shown in Figure.4

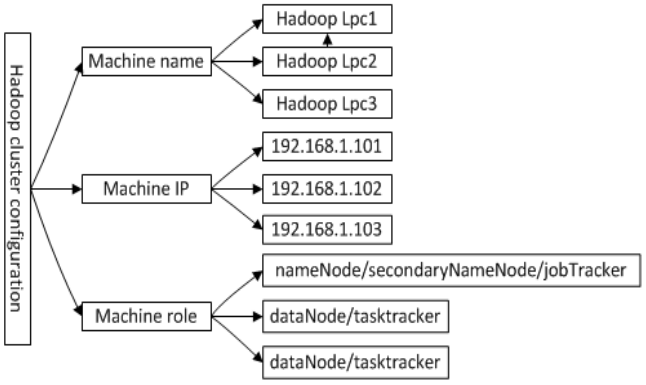


Figure 5. Build a cluster environment

B. Data Comparison Experiment

1) UCI experimental data

This experiment selects the retail file in the UCI database (association rules to study the classic data set) as the experimental transaction data set. By comparing the MEC-Apriori algorithm with the traditional Apriori algorithm, the results show that the time performance of the MEC-Apriori algorithm has been greatly improved in the mining of frequent itemsets and candidate itemsets, thus verifying the efficiency and feasibility of the improved algorithm.

2) Implementing the MEC-Apriori Algorithm Model

First, the experimental data set in the file retail is selected, and the data set in the retail is mined using the new MEC-Apriori algorithm, and then the association rule is obtained according to the user-defined support degree and the confidence threshold.

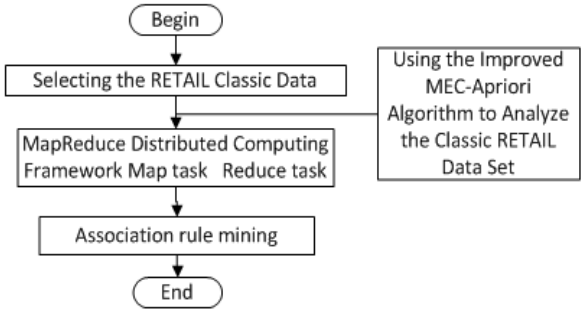


Figure 6. Simulation flow

3) Experiment content and result analysis

Experiment 1: Performance Comparison between Single Apriori Algorithm and MEC-Apriori Algorithm

The transaction data set for this experiment is stored as a file, Performance analysis of mining time before and after improved with 3 nodes Hadoop cluster test algorithm. First, on the premise that the number of nodes in the Hadoop cluster is unchanged, continuously increase the number of item sets in the experimental data item set, and set the minimum support to the same, that is, min_sup=0.3. The experimental results are shown in Table 1.1.

TABLE.I. COMPARING APRIORI WITH MEC-APRIORI MINING TIME

Transaction itemsets	Apriori Mining time/s	MEC-Apriori Mining time /s
2050	18.8	12.6
4150	25.4	14.8
6300	35.6	22.8
8150	59.2	35.7
11040	72.6	40.5

According to the experiment, the result obtained, convert the result to a line chart to make it more intuitive, Figure 4 shows the time Performance between MEC-Apriori and Apriori.

Horizontal axis: number of transaction item sets. Vertical axis: time/s.

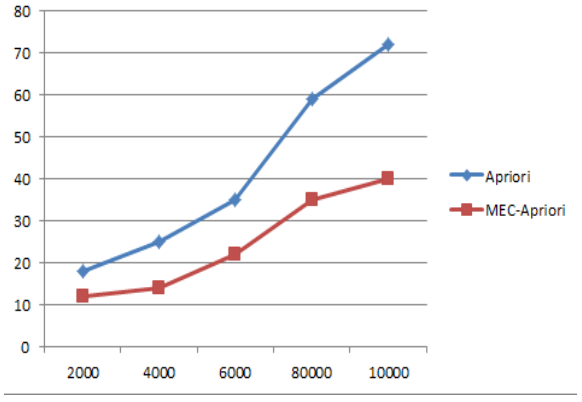


Figure 7. Improved and improved time performance charts

From the figure 4, the MEC-Apriori algorithm and the classical Apriori algorithm are on the premise of the same number of transaction itemsets, it is often better than Apriori algorithm in temporal performance, and with the increasing number of transaction item sets, apriori algorithm running on a computer can significantly improve the time of mining analysis. However, with the MEC-Apriori algorithm, as the number of transaction item sets increases, the time performance is getting better and better. Because with the increase in the number of transaction items, the nodes of the distributed cluster will gradually increase. In summary, the improved MEC-Apriori algorithm is superior to the classic Apriori algorithm in temporal performance.

Experiment 2: Performance Comparison between Apriori Algorithm and MEC-Apriori Algorithm under Different Supporting Degrees.

First ,this paper test the data set RETAIL, select the minimum support threshold range [0.02, 0.20]. And within this range, evenly increase the step: 0.02, so there will be a threshold of 10. Then, this paper use the data set retail to run the Apriori algorithm and the MEC-Apriori algorithm respectively, and record the running time (Note that the running time is second). Figure 5 shows the experimental data obtained by executing the above three algorithms. Horizontal axis: support; vertical axis: time/s.

Experiments show that the MEC-Apriori algorithm runs much less time than the Apriori algorithm under different support levels. The higher the support, the Apriori algorithm will run a little longer than the MEC-Apriori algorithm. In summary, the temporal performance of the MEC-Apriori algorithm under different support levels is always superior to the traditional Apriori algorithm.

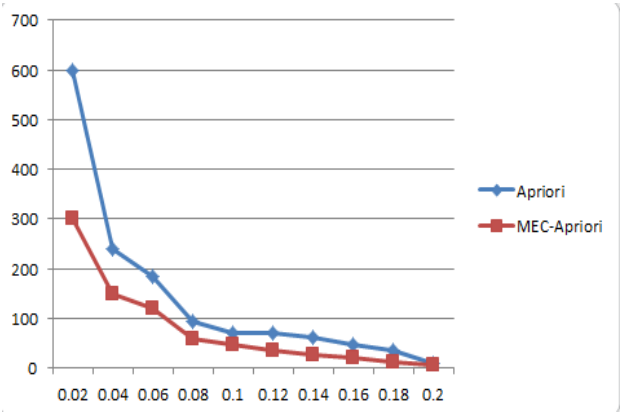


Figure 8. Performance comparison under different support levels

V. CONCLUSION

Aiming at the traditional Apriori algorithm, when mining frequent itemsets, you need to continuously scan transaction data sets , So that the system I / O overhead and other shortcomings. In this paper, we improved Apriori algorithm in three aspects: compression in the transaction, reducing the number of scanning areas, and simplifying the candidate set generation. At the same time, the improved algorithm is parallelized in the Hadoop framework. The simulation results show that compared with the traditional Apriori algorithm, the MEC-Apriori algorithm has good performance and security in temporal performance, mining frequent candidate itemsets and different support levels. However, it needs to be continuously improved in the future work.

REFERENCES

[1] K.WANG, Y.HE, J.HAN. Mining Frequent Itemsets Using Support Constraints. Proc2000 Int. Conf. Very Large Data Bases[J]. Cairo, Egypt, 2000.9: 43-52.

[2] Yan Xiaofei. Research on Association Rule Mining Algorithm[D]. Chongqing: Chongqing University, 2009:15-21.

[3] AGRAWAL R.SRIKANT R.Fast algorithm for mining a ssociation rules[C]//Proceedings of 20th Int. Conf. Very Large Data Bases(VLDB). Morgan Kaufman Press,1994:487-499.

[4] REN W J , YU B W. Improved Apriori Algorithm Based on Matrix Reducation[J]. Computer and Modern,2015,10. 2-3. (in Chinese)

[5] GUNARATHNE T , WU TL, QIU J ,et al .MapReduce in the Clouds for Science[C]//2010 IEEE Second International Conference on Cloud Computing Technology and Science (Cloudcom). IEEE,2010;565-572

[6] DEAN J,GHEMAWAT S. MapReduce: simplified data processing on large clusters[J]. Communications of the ACM, 2008, 51(1):107-113.

[7] HE B S, TAO M, YUAN X M. Alternating direction me-thod with Gaussian back substitution for Separable convex programming [J]. SIAM J. Optimization, 2012, 22(2): 313-340.

[8] HE B S,LIAO L Z,YUAN X M. Alternating projection based prediction-correction methods for structured variational inequalities[J]. Computational Mathematics, 2006, 24(6):693-710.

[9] CHEN Z M, WAN L, YANG Q Z. An Inexact Direction Methodfor Structured Variational Inequalities[J]. Journal of Optimization Theory & Applications, 2014, 163(2): 439-459.

[10] Lu Jiaheng. Hadoop Combat [M]. Beijing: Mechanical Industry Press, 2011: 17-128.

A Study of Intelligent Reading Model

Yu Jun
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, ShaanXi, China
e-mail: yujun@xatu.edu.cn

Hu Zhiyi
Engineering Design Institute
Army Academy of PLA
Beijing, 100000, China
e-mail: huzhiyi016v7@163.com

Kang Qinyu
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, ShaanXi, China
e-mail: 534739457@qq.com

Li Zhonghua
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, ShaanXi, China
e-mail: 761173763@qq.com

***Abstract*—In order to solve the problem of how to find out the required information quickly from a large number of reading texts, this paper constructs an intelligent reading model. The model adopts the principle of "the minority obeys the majority". The results of the classifier trained by the three algorithms, those are decision tree, Bagging and Gauss Bayes algorithm, are filtered to build an intelligent reading model. Based on the experimental results, the objective evaluation results of the new combinatorial algorithm are attained.**

Keywords-Natural language processing(NLP); Decision Tree; Bagging; Gaussian Bayesian Algorithm

I. INTRODUCTION

In recent years, with the rapid developing of the Internet and other emerging media, human beings have entered the era of information explosion. At the same time, more and more people hope that computers can understand human language so as to help human beings to perform various daily tasks better. Natural Language Processing(NLP)[1], as a typical example of artificial intelligence application in the practical field, is a necessary means for modern people to mine a large amount of data and information. Its main goal is to let computers learn to understand and use human natural language. Therefore, Natural Language Processing (NLP) has become a research hot spot in recent years.

At present, as one of the representative products of Natural Language Processing(NLP), "smart interactive technology [2]" has gradually penetrated into many products. However, many smart products can only recognize some specific commands. For example, when the input is "Open QQ (QQ is the abbreviation of Tencent QQ, which is an Internet-based instant messaging software developed by Tencent Company)", it can start QQ. But the input is "Look at QQ" and nothing happens. In addition, people have to read a lot of texts in daily life, such as novels, tutorials, etc. Sometimes you can solve the problem by just looking for a small part of the text without having to read through the whole article. For example, we can solve our legal doubts by looking for certain passages in the legal literature and be unnecessary to read the entire legal literature. Based on this, in order to make our reading more "intelligent", we need to establish an intelligent reading model that can use natural language to communicate with machines and let machines serve us in order to minimize the learning burden.

This paper builds an intelligent reading model. Since English is based on words, which are separated by spaces. However, Chinese is in the form of word, each of words in one Chinese sentence have to be connected for describing a complete meaning. For example, the English sentence "I am a student", Chinese means "I am a student". In English, the computer can easily know that "student" is a word by Spaces, However, in Chinese, "student" is made up of two words, which can only be combined to mean a word. Therefore, Chinese word segmentation is to divide the sequence of Chinese characters into meaningful words. Due to certain uncertainty in Chinese word segmentation, it is necessary to adopt many different technologies such as Jibe word segmentation [3] and TF-IDF weight algorithm [4]. In view of the simpleness of the previous model, a new combination model, namely our intelligent reading model, is constructed by combining multiple algorithms as well as adopting the principle of "the minority obeys the majority". Among them, the principle of "the minority obeys the majority" means that if a data is trained by three classifiers, the output result of classification is "0 / 0 / 1", Because there is a large number of 0 in results, so the data of the final result is 0. Finally, the validity of the model is verified by experiment and calculation.

II. THE OVERALL PROCESS OF BUILDING AN INTELLIGENT READING MODEL

The establishment of intelligent reading model includes five parts, these are data acquisition, data processing, feature extraction, training classifier and building model. The overall process is shown in figure 1. The details are as follows.

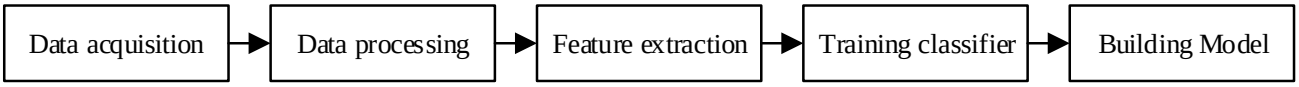


Figure 1. Overall framework

1) Data acquisition：Because there are all kinds of text on the network, and the quantity is huge, so the Python network crawler [5] is used to obtain the data. that is stored in a TXT file.

2) Data processing：Due to the inaccuracy of the word segmentation, it is necessary to use the algorithm based on information entropy to find new words, as well as put the new words into the custom dictionary, followed that process the text by Chinese word segmentation, stop word filtering and so on.

3) Feature extraction：For the processed data in step (2), the TF-IDF algorithm is used to extract the feature value and generate the word-text matrix.

4) Training classifier：Decision tree [6], Bagging [7] and Gaussian Bayesian algorithms [8] are used to train the word-text matrix generated by step (3).so that three classifiers could be obtained.

5) Building Model. For the three classifiers obtained in step (4), the principle of "the minority obeys the majority" is adopted to establish the intelligent reading model.

III. DESCRIPTION OF THE PROCESS OF MODELING

This paper constructs an intelligent reading model. Firstly, the Python network crawler is used for data acquisition. followed that Jieba word segmentation technology and TF-IDF weight algorithm were adopted to preprocess the sample data. Finally, Extracting the feature value, training classifier, establishing model and carrying out other operations. The detailed operation is described as follows.

A. Data acquisition

There are a variety of texts on the web. Due to the large number of data, Python web crawlers are usually used to obtain data. But some websites have anti-crawler mechanisms. Therefore, while designing a web crawler, a simulation operation of browser accessing is necessary. Through analyzing the web page source codes, regular expressions are used to obtain the required data. The library files, such as Beautifulsoup, Requests and Re can be used to crawl data. The content of the crawl is the problems and all their corresponding answers. Finally, the acquired data is stored in a TXT format file and is added a fixed tag name, so as to be convenient for later data processing.

B. Data processing

By analyzing the acquired data, a lot of noise in the text information can be found. For example, word segmentation is inaccurate, as well as there are a large number of stop words. If these noises are brought into the operation of word frequency statistics, it will not only reduce the processing speed, but also greatly affect the experimental results. Therefore, the first important thing is to preprocess the data. Data preprocessing is divided into three steps, which are generating and loading of the custom dictionaries, Chinese word segmentation and stop-word filtering. It is as shown in figure 2.

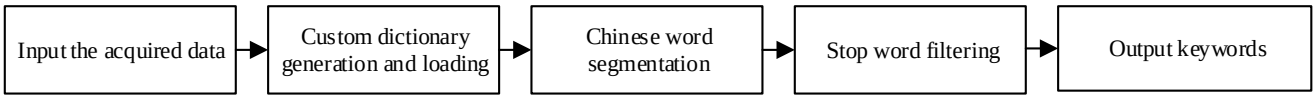


Figure 2. Flow of data preprocessing

1) Generation and loading of Custom Dictionary

Due to the number of words included in the Jieba dictionary is limited, which leads to the inaccuracy of text segmentation. For instance, it is inaccurate segmentation of people's names and place names. So that it is necessary to require a custom dictionary to improve the accuracy of word segmentation. The information entropy algorithm is used to find new words and generate custom dictionaries. After that, the generated custom dictionaries are loaded into the codes to improve the precision of word segmentation.

2) Chinese word segmentation.

After the above steps are completed, the word segmentation is beginning. This paper adopts a Chinese word segmentation module developed by Python-jieba word segmentation, which divides all data sets in Chinese. It combines rule-based and statistics-based methods [9].

The rule-based method means that, the word segmentation based on an existing dictionary which adopts manual rules such as forward maximum matching, backward maximum matching and bidirectional maximum matching. For example, for the sentence "Shanghai tap water comes from sea", forward maximum matching is used. It scans from forward to back, as well as makes the separated words exist in the dictionary and lets the words as long as possible. At last the sentence of "Shanghai / tap water / from / sea" can be obtained. This kind of method is simple and easy to implement, and the required data amount is not high.

The statistics-based method is to summarize the probability distribution of words and the common collocation between words from a large number of manually labeled corpus, and supervised learning is used to train the word segmentation model. For the sentence "Shanghai tap water comes from the sea", the most basic participle method

based on statistics is to try all possible participle schemes, Because of any two words, or need to cut, or without segmentation. For all possible word segmentation schemes, the probability of each scheme is counted according to the corpus, and then the one with the greatest probability is retained. Obviously, "Shanghai / tap water / come from / sea" is more likely than "Shanghai tap / water / come from / sea ", Because "Shanghai" and "tap water" appear more frequently in tagged corpus than "Shanghai tap" and "water". The result of partial participle is shown in figure 3.

100001	question: lane,occupy,Emergency,highway,How many	10000101	content: lane,occupy,Emergency,How many scores,Driving
100001	question: lane,occupy,Emergency,highway,How many	10000102	content: lane,driver,Emergency,confusion,highway
100001	question: lane,occupy,Emergency,highway,How many	10000103	content: lane,occupy,Emergency,Deduction,Illegal
100001	question: lane,occupy,Emergency,highway,How many	10000104	content: lane,parking,Emergency,The new traffic rules,occur
100001	question: lane,occupy,Emergency,highway,How many	10000105	content: The road traffic,Regulations,legislation,¥200,¥20
100001	question: lane,occupy,Emergency,highway,How many	10000106	content: lane,Emergency,2017,occupy,highway
100001	question: lane,occupy,Emergency,highway,How many	10000107	content: Emergency situations,lane,parking,Emergency,The new traffic rules
100001	question: lane,occupy,Emergency,highway,How many	10000108	content: lane,Emergency,occupy,2017,punishment
100002	question: The king of kung fu,Bride with white hair,play	10000201	content: The king of kung fu,Bride with white hair,Jet li,Jackie Chan
100002	question: The king of kung fu,Bride with white hair,play	10000202	content: Bride with white hair,kung fu,Jet li,Mars,The film

Figure 3. Participle screenshot

In figure 3, it can be found that there are a large number of meaningless modal particle in the results of word segmentation, which will greatly influence the final experimental results. Therefore, it is necessary to carry out the filtering of stop words.

3) Filter stop words

Stop Words [10] refers to words that appear frequently in the text without practical significance. Such as modal particle, adverbs, prepositions, conjunctions, etc. In order to save storage space and improve search efficiency, these meaningless stop words must be filtered out before processing text. To find out the stop word accurately, the following indicators can be used to measure the effectiveness of words.

a) Term Frequency. TF is a simple evaluation function whose value is the number of words occurring in the training set. The theoretical assumption of the TF evaluation function is that, when one word appears frequently in the text, it is generally regard as a noise word.

b) Document frequency. Similar to Term Frequency (TF), the theoretical assumption is that when one word appears frequently in the text, the word is generally regard as a noise word. The experimental result is shown as Table 1.

TABLE I. STOP WORD LIST(PARTIAL)

category	Stop Words
preposition	On, In, At, under, Beside, Behind, To, Over, with
pronoun	Everyone, everything, everywhere
...	...
adverbs	So, still, therefore, moreover, however

As shown in table 1, using filtered stop words to generate stop-word list, as well as to load the list into codes. Followed that the result of word is matched with the words in the stop-word list. If the matching is successful, the word from the result of the segmentation will be deleted.

C. *Feature extraction*

After the preprocessing of the above steps, although the stop word is removed, the sentence still contains a large number of words, which brings difficulties to the text vectorization process. Therefore, the main purpose of feature extraction is to minimize the number of words for being processed without changing the core content of the original text, so as to reducing the dimension of vector space, simplifying calculation and improving the speed and efficiency of text processing. Commonly used methods include term frequency-inverse document frequency(TF-IDF), information gain [11], X2 statistics, etc. Hereby TF-IDF algorithm is used to transform keyword information into weight vector in here. The steps are described as follows.

1) Calculating the word frequency, which is TF weight.

$$TF = \frac{\text{The number of times a word appears in the text}}{\text{The total number of words in the text}} \quad (1)$$

2) Calculating the inverse document frequency. that is IDF weight.

Firstly, a corpus is required to build up for simulating the language environment. The larger the IDF, the more concentrated this feature is in the text, it means that the more able the words are to distinguish the content of the text.

$$IDF = \log\left(\frac{\text{Total number of texts in a corpus}}{\text{Number of text containing the word} + 1}\right) \quad (2)$$

3) Calculating the Term Frequency Inverse Document Frequency (TF - IDF) values.

$$TF\text{-}IDF = TF \times IDF \quad (3)$$

The larger the TF-IDF value, the more important the word is. Calculating and sorting the TF-IDF value of each word in the text. The first six keywords of each question and corresponding answer in the text are found in turn, and the corresponding weight of the six keywords is returned. If there are less than 6 keywords, the residual weight is set to 0.

By using the TF-IDF algorithm, the text information is vectorized and the lexical text matrix is obtained. The details are described as follows.

$$\begin{array}{c} \quad d_1 \quad d_2 \quad K \quad d_n \\ \begin{array}{c} t_1 \\ t_2 \\ M \\ t_m \end{array} \begin{bmatrix} w_{11} & w_{12} & K & w_{1n} \\ w_{21} & w_{22} & K & w_{2n} \\ K & K & O & K \\ w_{m1} & w_{m2} & K & w_{mn} \end{bmatrix} \end{array}$$

Hereby, t_i (i=1,2,3..., n) is the feature item in document D. as well as w_{ij} (i, j=1,2,3..., n) is the weight of the feature item. The calculation formula is described as follows.

$$W_{ij} = TF\text{-}IDF = TF \times IDF \tag{4}$$

D. Training classifier

In this paper, the lexical - text matrix is trained with three algorithms of decision tree, Bagging and Gaussian Bayes. The specific steps of each algorithm are described below.

1) decision tree

Decision tree algorithm mainly includes feature selection and decision tree generation. The feature selection is based on the relationship between the information gain and the data set. According to the characteristics of the selected data set, the decision tree is generated recursively using ID3 algorithm [12]. The specific steps are described as follows.

a) Calculating information entropy.

In order to select the feature of good classification ability for training data, the information gain is introduced. And then, the calculation formula of information entropy is described as follows. Assume D is the training element group in the training set, its entropy can be expressed as follows.

$$info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \tag{5}$$

$$p_i = \frac{\text{Number of elements in this category}}{\text{Total number of training tuples}} \tag{6}$$

Hereby, m represents the total number of categories, and p_i represents the occurring probability of Category i which appears in the entire training tuple.

Entropy is a measure of the uncertainty of random variables. The actual meaning is the average amount of information required for the class label of Tuples D. The larger the entropy is, the greater the uncertainty of the variable. If the training tuple D is divided according to the characteristic attribute A, the expected information of D is described as formula (7). (Note: The expected information of D is conditional entropy, which is based on the classification of characteristic attribute A).

$$info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} info(D_j) \tag{7}$$

Dj is the classification of feature attribute A, as well as v is the number of types of the characteristic attribute A.

b) Calculating the information gain.

The information gain is the difference between the two information entropy.

$$gain(A) = info(D) - info_A(D) \tag{8}$$

As the above formula (8) shows, gain(A) represents the amount of information obtained by classifying A as a node. The more information, the more important A is.

c) ID3 algorithm is used to establish each child node in the tree. According to the characteristics of the data set selected by the information gain, the algorithm selects the feature with the maximum information gain as the judgment node and acts as the sub-node in the tree.

d) Using recursive thinking, repeat above steps from (1) to (3) so as to establish the decision tree.

2) Bagging integrated decision tree

Bagging is a technology of repeated sampling from data according to uniform probability distribution. The algorithm does use the different training set to fit a single member classifier in the ensemble classifier, as well as Bootstrap sampling is used by training set in the fitting process. which is a random sampling with a rewind. So bagging can improve the accuracy of unstable model, and reduce the degree of over-fitting. The final result of the algorithm is to construct a series of prediction functions, and combining them into a prediction function by voting. The process is shown in figure 4.

The steps are described as follows.

a) The bootstrap [13] method is used to select n training samples from the sample set, and using it as the training set T1-Tn. This process is executing for K times, and k subsets $\{T_1, T_2 ... T_K\}$ are selected.

b) K sample subsets are trained on their own training data on all attributes. And then k classification models are obtained.

c) According to the classification model obtained by the above steps, the value of each {P1, P2 …, Pk} model is predicted respectively.

d) The value $\{P_1, P_2, \ldots, P_k\}$ of each model is combined by the average method. The final result is output. The formula of the averaging method is described as follows.

$$P(x) = \frac{1}{K} \sum_{i=1}^{K} p_i(x) \tag{9}$$

Hereby, pi is the value of a model. K is the number of samples training.

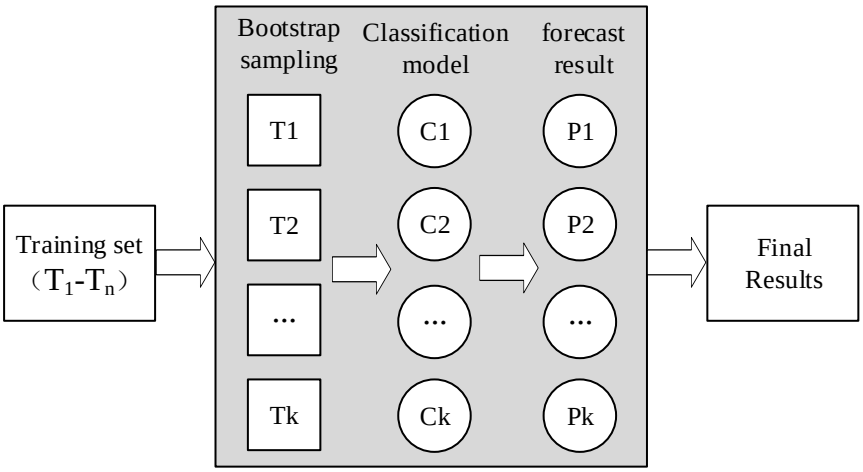


Figure 4. Bagging process

3) Gaussian Bayes algorithm

Compared with decision tree and Bagging algorithms, the greatest advantage is that when the large scale training set is selected, the Gauss Bayes algorithm only has relatively small number of features for each item, and the training and classification of the project is only a mathematical operation for the characteristic probability. Therefore, Gaussian Bayes algorithm has a fast speed when training large amount of data. The flow of this algorithm is shown as in Figure 5.

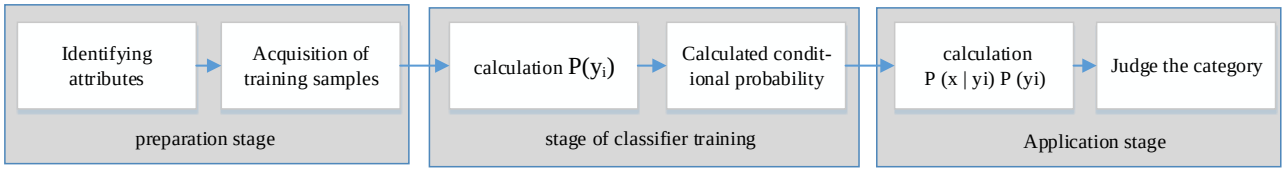


Figure 5. Gaussian Bayesian algorithm flow

As shown in figure 5, the entire algorithm flow can be divided into three phases.

The first is Preparation stage. This stage determines the characteristic attributes according to the specific situation, and dividing each feature attribute appropriately. And then, some of the items is classified by manually so as to form a training sample set. The input of this phase is all the data to be classified, and the output is the feature attribute and the training sample. This stage is the only stage that needs to be completed manually in the whole naive Bayesian classification. The quality of the classifier will have an important impact on the whole process, as well as be affected by the feature attributes, the classification of the feature attributes and the training samples.

The second is Classifier training stage. This stage is generating the classifier. Firstly, the occurrence frequency of each class is calculated in the training sample. And then the conditional probability of each category on the characteristic attribute is calculated. Finally, the results are recorded. The input is the feature attribute and the training sample, and the output is the classifier.

The third is Application stage. The task at this stage is to classify the classified items by using the classifier. The input of this stage is classifier and the item to be classified, and the output is the mapping relationship between items to be classified and categories. The specific steps are described as follows.

The specific steps are described as follows.

a) Assume $x = \{x_1, x_2, ..., x_m\}$ be an item to be classified, and each x_i be a characteristic attribute of x.

b) Setting have a set of y categories where $y_1=0, y_2=1$.

c) Calculating conditional probability $P(x_i \mid y_j)$ .The formula of (10) is shown as follows.

d) If it is existed as $P(x \mid y_k) = \max\{P(x_i \mid y_j)\}$, then it will be $x \in y_k$.

e) According to Bayesian theorem, the following formulas can be obtained. The formula of (11) is shown as follows.

f) The class that maximizes the value of P(x|yi) P(yi) is found out, and the items to be classified fall into this category.

$$P(x_i \mid y_j) = \{P(x_1 \mid y_1), P(x_2 \mid y_1), ..., P(x_m \mid y_1), P(x_1 \mid y_2), P(x_2 \mid y_2), ..., P(x_m \mid y_2)\} \quad (10)$$

$$P(x \mid y_i)P(y_i) = P(x_1 \mid y_j)P(x_2 \mid y_j)...P(x_m \mid y_j)P(y_j) = P(y_i)\prod_{i=1}^{m} P(x_i \mid y_j) \quad (11)$$

E. Building Model

The above three algorithms is adopted to construct the three model. Followed that, the principle of "the minority is subordinate to the majority " is used to reconstruct a new model. The model is called as intelligent reading model. The principle of "the minority is subordinate to the majority" means that, if a data is trained by three model, the output result of classification is "0 / 0 / 1". Because of the number of '0' in the result is more than the number of '1', the final result of the data is 0.

IV. VERIFICATION AND ANALYSIS OF THE MODEL

The data from the testing set are input into the intelligent reading model, as well as the processed results analyzed. The quality of the model is measured by two technical indicators, that is, Accuracy and F-Measure value.

The two-dimensional confusion matrix is shown in Table 2. The meaning of " the forecast is wrong, the actual is wrong (TN)" is that, the actual label category of the data is wrong, and it is still wrong after prediction. Based on the two-dimensional confusion matrix shown in Table 2, the formula (12) of the accuracy rate and the formula (13) of F-Measure are given as below.

TABLE II. TWO-DIMENSIONAL CONFUSION MATRIX

		actual value	
		positive	**Wrong**
Forecast value Such as type (10)	**positive**	The forecast is positive, The actual is positive (TP)	The forecast is positive, the actual is wrong (FP)
	Wrong	The forecast is wrong, the actual is positive (FN)	The forecast is wrong, the actual is wrong (TN)

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \quad (12)$$

$$\text{F-Measure} = \frac{2*TP}{2*TP + FP + FN} \quad (13)$$

As shown in formula (12), the accuracy rate refers to the proportion of successful data in all predicted data. The predicted success means "the predicted value is same as the actual value". It includes two kinds of labels such as "the forecast is positive, the actual is positive (TP)" and "the forecast is wrong, the actual is wrong (TN)". When users ask some questions, they only want the right answers. Therefore, TN label is not necessary. As shown in formula (13), The F-Measure value is a comprehensive evaluation index of accuracy rate and recall rate. Because it does not include TN label, it is often used to evaluate the classification model.

Accuracy is a very objective evaluation index, but sometimes the accuracy rate does not represent the quality of the algorithm. Especially in the case of imbalance of positive and negative samples, the accuracy evaluation index has great defects. The most common F-Measure method is the weighted harmonic average of the accuracy rate and recall rate (the recall rate is the measure of the cover surface). Because the F-Measure method comprehensively considers the accuracy rate and recall rate, it effectively avoids the problem of unbalanced data distribution. Therefore, comparing with the accuracy rate, the f-measure method can better reflect the quality of the algorithm. Among them, the higher the value of F-Measure, indicating the better classification results of the corresponding algorithm. If we combine the three algorithms such as decision tree, Bagging and Gauss Bayes, the results of the combination algorithm and the single algorithm are shown in Table 3.

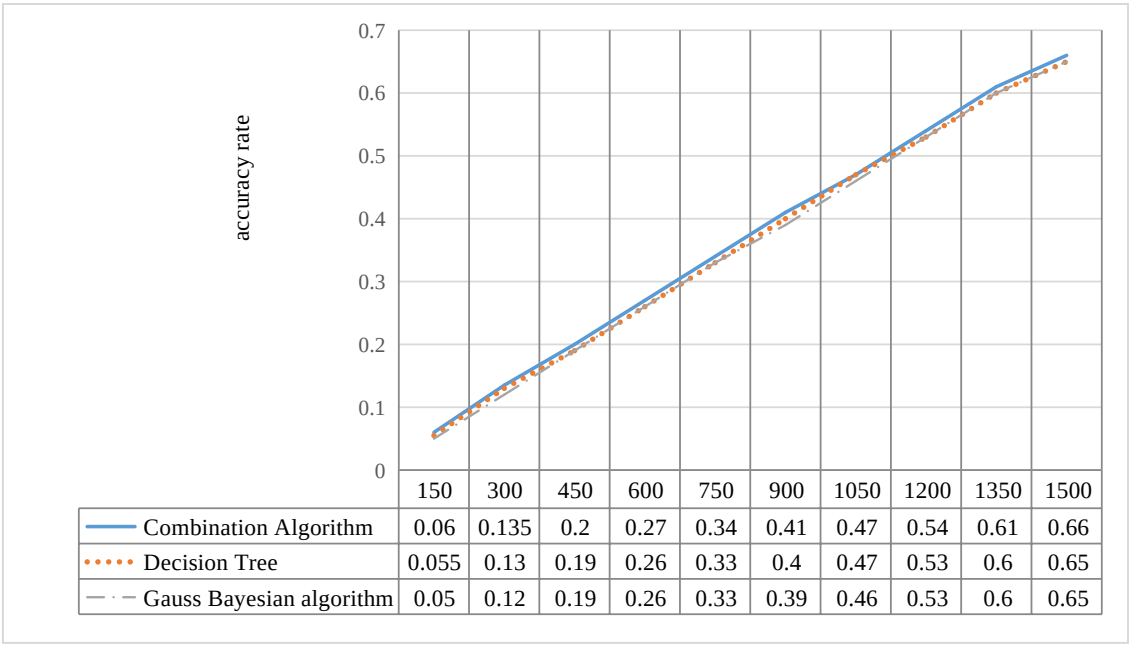
TABLE III. COMPARISON OF PREDICTION RESULTS

	decision tree	**Bagging**	**Gauss Bayesian**	**Combination algorithm**
Accuracy	0.6569	0.7105	0.7093	0.7112
F-Measure	0.3354	0.1933	0.2504	0.3381

Table 3 shows that the accuracy of the combination algorithm is 0.7112, while the other three separate algorithm accuracy is 0.6569, 0.6569 and 0.7105. Therefore, the combination algorithm is more accurate than the other three methods. In addition, the F-Measure value of the combined algorithm is 0.3381. The F-Measure value of the other three methods is 0.3354, 0.1933 and 0.2504. Therefore, the combination algorithm is better than the other three separate algorithms in terms of F-Measure.

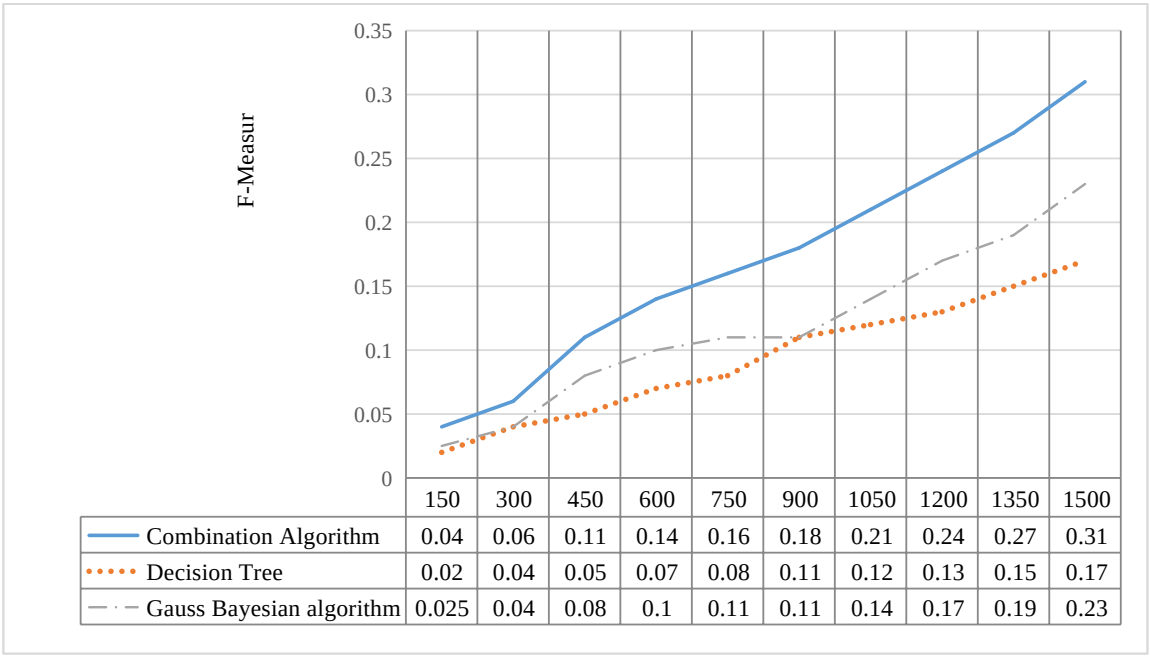
Whether it's accuracy or F-measure, the result of the combined algorithm is better than that of the other three separate algorithms. Therefore, the intelligent reading model is based on the combination of decision tree, Bagging and Gauss Bayesian algorithms.

In order to verify the superiority of the method, we are selecting about 8000 pieces of data for experimental verification. The experimental results are shown in figure 5 and figure 6.



	150	300	450	600	750	900	1050	1200	1350	1500
Combination Algorithm	0.06	0.135	0.2	0.27	0.34	0.41	0.47	0.54	0.61	0.66
Decision Tree	0.055	0.13	0.19	0.26	0.33	0.4	0.47	0.53	0.6	0.65
Gauss Bayesian algorithm	0.05	0.12	0.19	0.26	0.33	0.39	0.46	0.53	0.6	0.65

Figure 6. Accuracy comparison diagram



	150	300	450	600	750	900	1050	1200	1350	1500
Combination Algorithm	0.04	0.06	0.11	0.14	0.16	0.18	0.21	0.24	0.27	0.31
Decision Tree	0.02	0.04	0.05	0.07	0.08	0.11	0.12	0.13	0.15	0.17
Gauss Bayesian algorithm	0.025	0.04	0.08	0.1	0.11	0.11	0.14	0.17	0.19	0.23

Figure 7. F-Measure comparison

Experiment in figure 6, the X-axis represents the amount of data in the experiment, and the Y-axis represents accuracy, which shows that comparing with the three algorithms of decision tree, Bagging and Gauss Bayes, the combination algorithm has a slightly better accuracy. in figure 7, the X-axis represents the amount of data, and the Y-axis represents the F-Measure. The F-Measure of the combined algorithm is obviously superior to other three separate algorithms.

V. CONCLUSION

The reading model constructed in this paper makes reading more intelligent. Aiming at the problem of natural language input, the corresponding answer can be given according to the existing TXT content. According to the experimental data, it can be concluded that the intelligent reading model based on the combination of decision tree, Bagging and Gauss Bayesian algorithm has a good classification ability.

REFERENCES

[1] Xi Xuefeng, Zhou Guodong. A study of Deep Learning for Natural language processing [J]. Journal of Automation! 42 (10): 1445-146

[2] Chen Lian. Research and Application of key Interactive Technology based on Web Intelligent Education platform [D]. Graduate School of Chinese Academy of Sciences (Chengdu Institute of computer Application).

[3] Han Dongxu, Chang Baobao. Domain adaptability method of Chinese word Segmentation Model [J]. Journal of computer Science, China (02): 272-281.

[4] Yang Bin, Han Qingwen, Lei Min, Zhang Yapeng, Liu Xiangguo, Yang Yaqiang, Ma Xuefeng. Short text Classification algorithm based on improved TF-IDF weight [J]. Journal of Chongqing University of Technology (Natural Science) 30 (12): 108-113.

[5] Qian Cheng, Yang Xiaolan, Zhu Fuxi. Python-based web crawler technology [J]. Heilongjiang Science and Technology Information 2016 (36): 273.

[6] Wang Daoming, Lu Changhua, Jiang Weiwei, Xiao Mingxia, Li inevitable. Research on decision Tree SVM multiple Classification method based on Particle Swarm Optimization algorithm [J]. Journal of Electronic Measurement and Instruments 29 (04): 611-615.

[7] Bi Kai, Wang Xiaodan, Yao Xu, Zhou Jindeng. An adaptive selective integration based on bagging and confusion matrix [J]. Chinese Journal of Electronic Science (EJ). 42 (04): 711-716.

[8] Zhu Mingmin. Study on Bayesian Network Structure Learning and reasoning [D]. Xi'an University of Electronic Science and Technology.

[9] Zan Hongying, Zuo Weisong, Zhang Kunli, Wu Yunfang. A study on emotion Analysis combined with rules and Statistics [J]. Computer Engineering and Science 33 (05): 146-150.

[10] Gu Yijun, Fan Xiaozhong, Wang Jianhua, Wang Tao, Huang Weijin. Automatic selection of Chinese stops word list [J]. Beijing Institute of Technology Proceedings 2005 (04): 337-340.

[11] Liu Qinghe, Liang Zhengyou. A feature selection method based on information gain [J]. Computer Engineering and applications 47 (12): 130-132 + 136.

[12] Huang Yuda, Fan Taihua. Analysis and Optimization of decision Tree ID3 algorithm [J]. Computer Engineering and Design! 33 (08): 3089-3093.

[13] Liu Jian, Wu Yi, Tan Lu. Improvement of bootstrap method for self-help sampling [J]. Mathematical Theory and Application 2006 (01): 69-72.